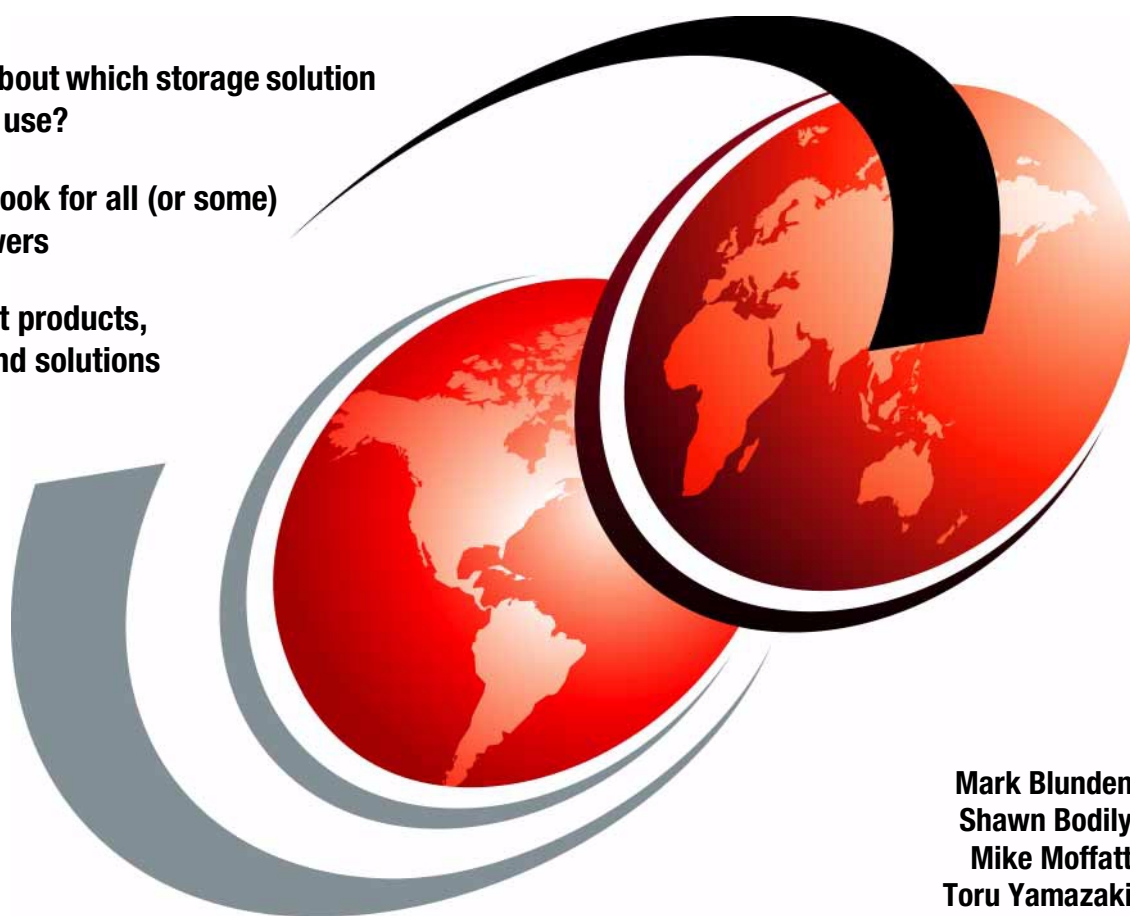IBM

Enterprise Storage Solutions Handbook

Confused about which storage solution
you should use?

Read this book for all (or some)
of the answers

Learn about products,
features, and solutions

Mark Blunden
Shawn Bodily
Mike Moffatt
Toru Yamazaki

Redbooks

IBM International Technical Support Organization

Enterprise Storage Solutions handbook

January 2001

> **Take Note!**
>
> Before using this information and the product it supports, be sure to read the general information in Appendix B, "Special notices" on page 383.

Third Edition (January 2001)

This edition applies to products and services available at the time of writing across a variety of platforms as described in this redbook.

Comments may be addressed to:
IBM Corporation, International Technical Support Organization
Dept. 471F Building 80-E2
650 Harry Road
San Jose, California 95120-6099

When you send information to IBM, you grant IBM a non-exclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

Contents

Figures

Tables

Preface

Mission-critical data to run your e-business is being distributed from enterprise servers to functional UNIX servers, departmental Windows NT servers, and even your desktop PCs, to get the maximum benefit out of each platform. To avoid data loss, the storage resources holding that data must be well managed. IBM mainframes such as S/390 systems have offered system-managed storage (SMS) capability.

This IBM Redbook is intended to help you understand what enterprise storage management is; the key technologies to know; the IBM subsystems, software, and solutions available today; and how to implement various enterprise storage administration tasks.

Introduction to Enterprise Storage Management first discusses the importance of storage management and contemporary storage management challenges being faced by storage administrators. It then introduces the Seascape architecture, Storage Area Network (SAN) Initiative, and Data Sharing solutions.

Understanding Storage Subsystems discusses key technologies used in storage subsystems today, in order to better understand the characteristics of those subsystems. It then introduces the IBM storage subsystems that are currently available for both Open Systems and S/390 platforms.

Understanding Storage Management Software again discusses key technologies used in storage management software today, and provides overviews of the IBM storage management software, including Tivoli Storage Manager, DFSMS/MVS, and StorWatch.

Implementing Enterprise Storage Administration focuses on how to implement your own enterprise storage administration. We have extended the tasks originally developed for S/390 systems to the heterogeneous environments in use today. We also cover various IBM services that you can exploit to establish your enterprise storage management environment much faster.

We hope this book will be of help to you in establishing your own enterprise storage management environment.

Note: We have changed the former title of this book from *Principles of Enterprise Storage Management* to *Enterprise Storage Solutions Handbook* to better describe its contents.

The team that wrote this redbook

This redbook was produced by a team of specialists from around the world working at the International Technical Support Organization San Jose Center (Figure 1).

Mark Blunden is the Project Leader for Storage Systems Disk at the International Technical Support Organization, San Jose Center. He has co-authored ten previous redbooks and co-managed seven other redbooks. Mark also teaches IBM classes worldwide on all areas of storage. He has worked for IBM for 20 years in many areas of the IT business. Before joining the ITSO in 1998, Mark worked in Sydney, Australia, as an Advisory Storage Specialist.

Shawn Bodily is an HACMP/CATE Specialist working on the Advanced Technical Support (ATS) pre-sales team located in Dallas, Texas. Before joining ATS in March 2000, he was the team leader of AIX Support Line's HA_DASD team. He has 6 years of AIX experience, with the last 3 years specializing in High Availability and Disk Storage Subsystems.

Mike Moffatt is an IT Specialist in the Americas Storage Techline in Toronto. He provides technical and solution pre-sales support to IBM Storage Specialists and Business Partners. Before joining Techline in 1998, he installed and relocated IBM's large systems hardware with the Complex Systems Install Team. Mike has 19 years of experience as a storage hardware specialist providing customer service, support, and project implementation.

Toru Yamazaki is an Advisory IT Specialist working as a team leader for Storage Systems at Asia Pacific Advanced Technical Support, IBM Japan. He brought his original idea of this book to the International Technical Support Organization when he was a project leader in 1998 and has conducted the first two editions. His broad range of technical support activities include consultation on enterprise storage management implementation and performance benchmarking on competitive selling situations, as well as presentations at international conferences.

The authors of previous editions of this redbook are:

- Jose Miguel Casas, Distrilogie Iberica
- Daniel Asselin, IBM France
- Masaaki Ida, IBM Japan
- Marius Muller, IBM France

- Jacqueline Tout, IBM Australia

Thanks to the following people for their invaluable contributions to this project:

- Alice Bird, IBM Storage Systems Group
- Bob Friske, IBM Storage Systems Group
- John Humphrey, IBM Storage Systems Group
- Wilbur Jackson, IBM Storage Systems Group
- John Tyrrell, IBM Storage Systems Group
- Craig Welch, IBM Storage Systems Group
- Freddy Saldana, Tivoli Systems
- Ricaldo Haragutchi, IBM Cary
- Doug Mentzer, IBM Tucson
- Sergei Varbanov, IBM Rolling Meadows
- Jon Tate, ITSO, San Jose
- Roland Leins, ITSO, San Jose
- Charlotte Brooks, ITSO, San Jose
- Mary Lovelace, IBM Learning Services, San Jose
- Emma Jacobs, ITSO, San Jose
- Yvonne Lyon, ITSO, San Jose
- Claudia Traver, ITSO, San Jose

Comments welcome

Your comments are important to us!

We want our Redbooks to be as helpful as possible. Please send us your comments about this or other Redbooks in one of the following ways:

- Fax the evaluation form found in "IBM Redbooks review" on page 413 to the fax number shown on the form.
- Use the online evaluation form found at **ibm.com**/redbooks
- Send your comments in an Internet note to redbook@us.ibm.com

Figure 1. The team that wrote this book: Shawn, Mike, Mark, and Toru

Part 1. Introduction to Enterprise Storage Management

Chapter 1. Enterprise Storage Management overview

In this chapter, we discuss both the current challenges facing storage administrators, and also some of the historical background to describe how we have arrived where we are.

1.1 Storage as a computing resource

Computing resources basically consist of three major components:

- Processor resource
- Storage resource
- Network resource

This section first shows the connection types of these resources, and then discusses the importance of storage resource management.

1.1.1 Connection types

Local storage — The traditional connection type of storage resource is local storage (see Figure 2). Storage resource is connected only to the local host and is not physically shared. Data stored on local storage is primarily used by the local host.

Bare hard disk drives and tape drives can be configured as local storage. These component devices are introduced in 6.2, "Standards, technologies, and building blocks" on page 80, and are detailed in Chapter 7, "Disk storage subsystems" on page 121 and Chapter 8, "Removable storage solutions" on page 153.

To share data with other systems, the following hardware and software data sharing solutions can be used: SAN Data Gateway (see 9.1, "SAN Data Gateway for serial disk" on page 185, and 9.3, "SAN Data Gateway for UltraSCSI disk and tape storage" on page 189).
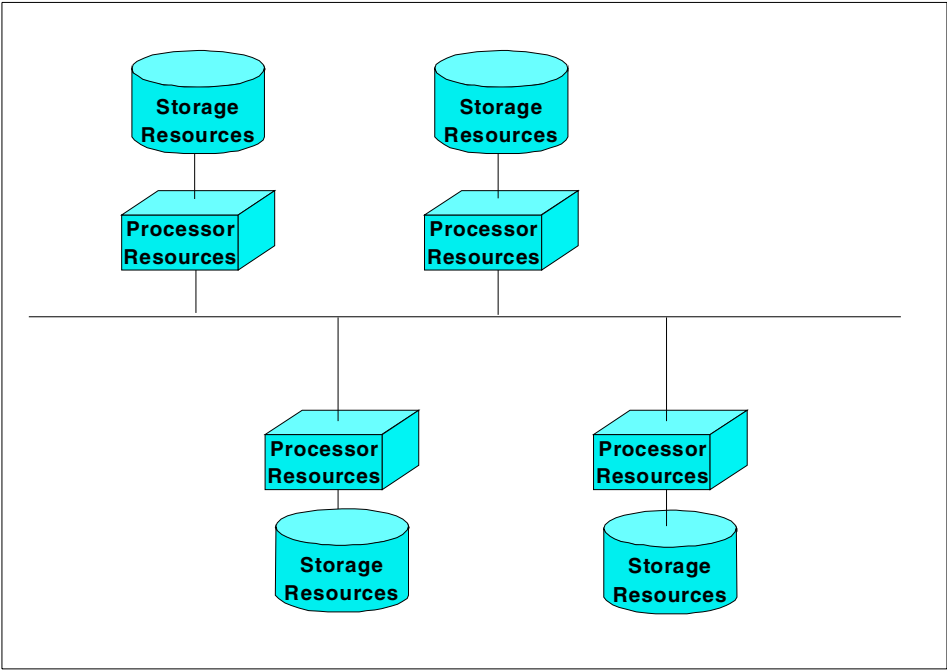
Figure 2. Local storage

- **Network attached storage** — The storage resource expected to be shared among multiple processor resources is attached directly to the network (see Figure 3). This second connection type is called *network-attached storage*.

  3995 Optical Library Data Server and 3466 Network Storage Manager are typical examples that support this connection type. For a detailed description of each product, see 8.8, "3995 Optical Library Data Server" on page 181; and 10.1, "IBM 3466 Network Storage Manager" on page 205.
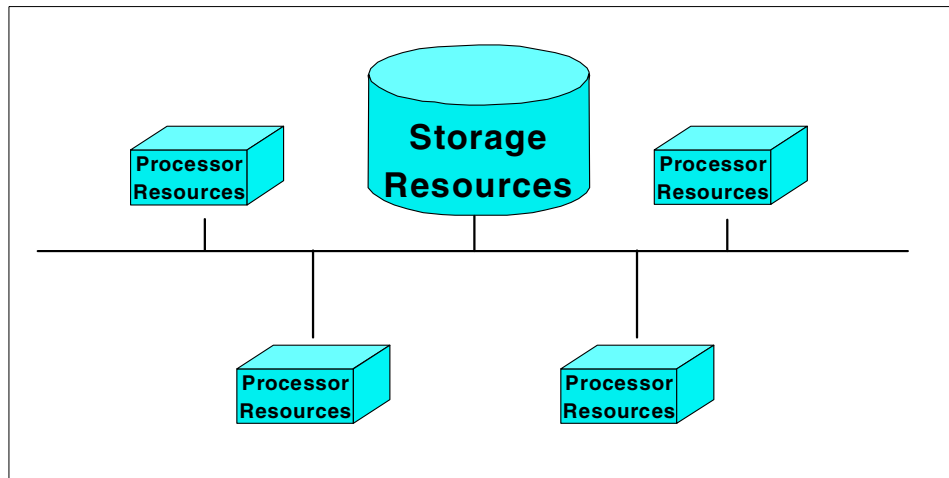
Figure 3. Network attached storage

- **Storage Area Network** — The Storage Area Network (SAN) is a new concept of storage resource connection. In this manner, the storage resource is directly connected to multiple hosts without using network resources to avoid conflicts between I/O requested data and application data (see Figure 4, and refer to Chapter 3, "Storage Area Network Initiative" on page 19).

  Enterprise Storage Server (ESS) and RAMAC Virtual Array (RVA) support this connection type. VSS is used for heterogeneous open system platforms (see 7.1, "The IBM Enterprise Storage Server (ESS)" on page 121).
  RVA is used for homogeneous S/390 systems (see 7.5, "RAMAC Virtual Array Turbo Model X83" on page 144).
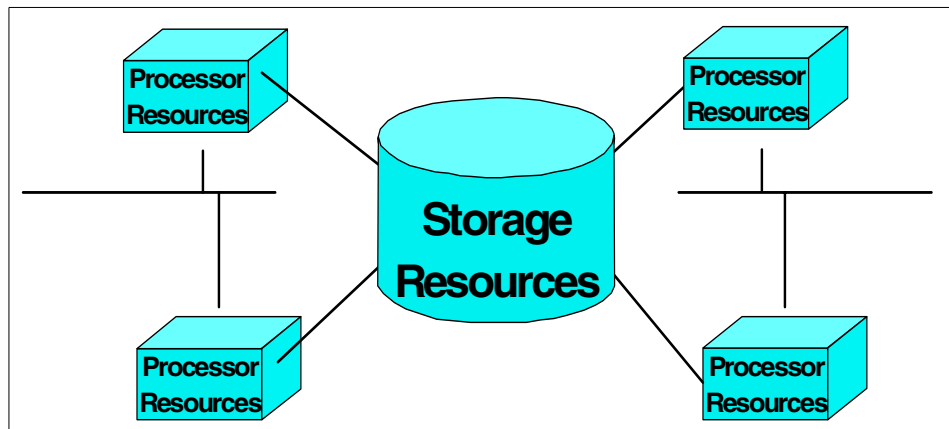
Figure 4. Storage Area Network

Other IBM storage subsystems and solutions described in subsequent parts of this book support various types of connections to optimize your storage configuration. See Part 2, "Understanding storage subsystems" on page 71, and Part 3, "Understanding storage management software" on page 211 for more details.

1.1.2 Importance of storage resource management

We take a look at the importance of storage resource management in terms of capacity, performance, and availability.

- **Capacity** — Managing a certain amount of free storage capacity is important in order to avoid failing your application. If you do not have enough capacity to store your data, programs will immediately fail. In the case of processor or network capacity shortage, for example, even if your

processor resource hits 100% usage, you can still wait for completion without any recovery action.

The hierarchical storage management (HSM) function provided by both Tivoli Storage Management and Data Facility Storage Management Subsystem/MVS (DFSMS/MVS) automatically manages free space based on the policy that the storage administrator sets. HSM basics are described in 11.2.6, "Hierarchical Storage Management" on page 216. Product specific implementation is also discussed in 12.3, "Space management" on page 233, and in 13.3, "Space management" on page 260.

- **Performance** — Only the storage resource component involves mechanical movements such as moving heads and spinning disks. Thus, storage speed is much slower than for the processor resources and even the campus network resources. Today, many storage subsystems have buffers, called *cache*, to compensate for performance gap.

  Better managing the performance of storage resource provides greater impact on overall system performance than can be obtained by optimizing any other system component.

  DFSMS/MVS automatically manages performance, and StorWatch Reporter is used to monitor your enterprise storage performance.

- **Availability** — Basic recovery procedures for processor resource and network resource failure involve simply replacing equipment. This is because these resources deal only with in-flight data.

  On the other hand, if something happens on storage resources, in addition to equipment replacements, you must take time to recover data. Latest technologies such as data mirroring and RAID help protect your data to a certain degree but are still not efficient for errors caused by people or software. Thus, availability of storage resource is vitally important.

1.2 Historical perspectives on storage management

In this section, we take a look at what IBM has accomplished in storage management. The following are some of the firsts in IBM's history of leadership:

1978 **RAID** (Redundant Arrays of Independent Disks) technology, an area where IBM has always led. In 1978 IBM received the first patent for a disk array subsystem. IBM subsequently co-sponsored the research by the University of California at Berkeley that led to the initial definition of RAID levels in 1987.

1978   Hierarchical Storage Manager (**HSM**), providing customers system-delivered migration of inactive data from disk to less expensive storage media.

1984   The first thin-film, 18-track head in a tape drive, setting a new industry standard. IBM also introduced chromium dioxide in place of iron oxide for tape media, permitting the highest density in the industry. "High Speed Locate" was another first in the tape industry. Electronic buffers replaced vacuum columns in tape drives, helping to increase data rate to three megabytes per second.

1988   The first full-function, automatic environment for management of storage devices, Data Facility Storage Management Subsystem (**DFSMS**).

1993   The **ADSM** family (nowadays known as Tivoli Storage Management), client/server storage management solutions for the multi-platform multi-vendor data residing on most networks. Backed-up data can be stored on AIX, AS/400, OS/2, MVS, VM or VSE servers.

1998   **StorWatch**, an enterprise storage resource management software, was announced.

See `http://www.storage.ibm.com/firsts/timeline.htm` for more details.

1.3 Contemporary storage management challenges

The following sections summarize some of the contemporary storage management challenges being faced by storage administrators today.

1.3.1 Network computing

With the growing reliance on network computing and the widespread use of distributed systems throughout the enterprise, the storage administration tasks has become increasingly difficult.

The StorWatch product family helps storage administrators centralize enterprise storage resource management. See Chapter 14, "StorWatch" on page 287 for details.

1.3.2 Data sharing

With today's increasingly sophisticated applications, such as business intelligence and e-business, information must be accessible to anyone who needs it.

InfoSpeed provides an optimal, one-step, data-copy sharing approach to meet today's challenges of open connectivity, flexibility, and performance.

1.3.3 Server consolidation

Server consolidation is a way to centralize business computing workloads to reduce cost, complexity, and management overhead and, in general, to free up business professionals to focus on their business instead of IT management tasks.

With the complexities of server consolidation and new enterprise application development, it has become very difficult to plan separate disk storage capacity requirements for S/390 and other platforms.

Enterprise Storage Server can be attached to multiple open platforms (open systems and S/390) to consolidate storage subsystems. See 7.1, "The IBM Enterprise Storage Server (ESS)" on page 121 for details.

SAN is also a key player in the server consolidation allowing fast communication in the connected world. See Chapter 3, "Storage Area Network Initiative" on page 19 for details.

1.3.4 Continuous availability

Customers need to have data accessible seven days a week and 24 hours a day. This diminishes the time available for performing backups, and doing housekeeping tasks. It also reduces the time allowable for data to be recovered in the event of a failure that renders the data inaccessible.

Historically, IT has focused on business continuance solutions for centralized, OS/390 applications and systems. The explosion of distributed UNIX servers, and their growing requirement to support daily business activities, demand the same level of business continuance support originally developed for OS/390. The survival of business enterprises depends upon IT providing business continuance support enterprise-wide.

In order to control and make their storage management efforts more efficient and to better manage their storage investments, customers should think about consolidating their storage management skilled resources. Today, storage solutions make that consolidation easier. See Chapter 16, "Enterprise storage administration" on page 305 for more details.

IBM storage management software and several storage subsystems provide availability management functions to achieve various needs from daily incremental backup to duplicating data to disaster recovery. You can see these functions throughout this book.

1.3.5 e-business

The Web is changing every aspect of our lives, but no area is undergoing as rapid and significant a change as the way businesses operate. As businesses incorporate Internet technology into their core business processes, they start to achieve real business value.

Today, companies large and small are using the Web to communicate with their partners, to connect with their back-end data-systems, and to transact commerce. This is e-business.

For further information about e-business and how IBM can help your current business processes to improve operating efficiencies, visit the IBM Web site:

    www.ibm.com/e-business

1.3.6 Business Intelligence

Business Intelligence (BI) is an exciting new concept in the IT industry. It employs a range of technologies (both old and new) designed to help businesses gain a competitive advantage. Most businesses have their data locked up in many different discrete databases and files that are not useful outside a specific application.

The range of new technologies in BI changes all that, making data accessible and easy to use for a company's vital strategic planning and decision making. Two of the more mature technologies in use for BI are data warehousing and decision support systems (DSS). These have been around since the early 1980s.

The best-established BI technologies are the following:

- **Data warehousing (DWH)**

  An information warehouse or data warehouse consists of data which has been extracted from operational systems, cleansed and transformed, and then loaded into a relational database where it can be easily accessed. Information about the data — metadata — is then used to create a catalog or directory so users can find out what data exists and what types of queries can be formulated.

- **Decision support systems (DSSs)**

  The purpose of a decision support system (DSS) is to enable users to predict the impact of business decisions before they are made. A DSS is an integrated set of programs which allows businesses to analyze the information stored in a data warehouse or data mart.

- **Online Analytical Processing (OLAP)**

  This technology allows users to analyze information that has been summarized into multidimensional formats. OLAP engines store data in relational databases and use SQL to extract the data as needed. OLAP tools are typically used to perform trend analysis on sales and financial information. For example, they enable users to drill down into masses of sales statistics in order to pinpoint products which exhibit specific levels of activity.

- **Multi-Dimensional Databases (MDDBs)**

  These databases, also known as Multidimensional OLAP or MOLAP, offer faster versions of traditional OLAP products because data consolidation is done ahead of time, prior to user queries.

- **Data mining**

  Data mining is one of the newest and most sophisticated Business Intelligence technologies. It can provide businesses with insights that are otherwise impossible to obtain. For example, while OLAP tools allow a company to compare sales revenues for two quarters, data mining technology would enable a comprehensive search through all sales data and then present the user with a hypothesis.

- **Data marts**

  A data mart is a small information subset of the data in a data warehouse. A data warehouse can contain numerous data marts (see Figure 5).
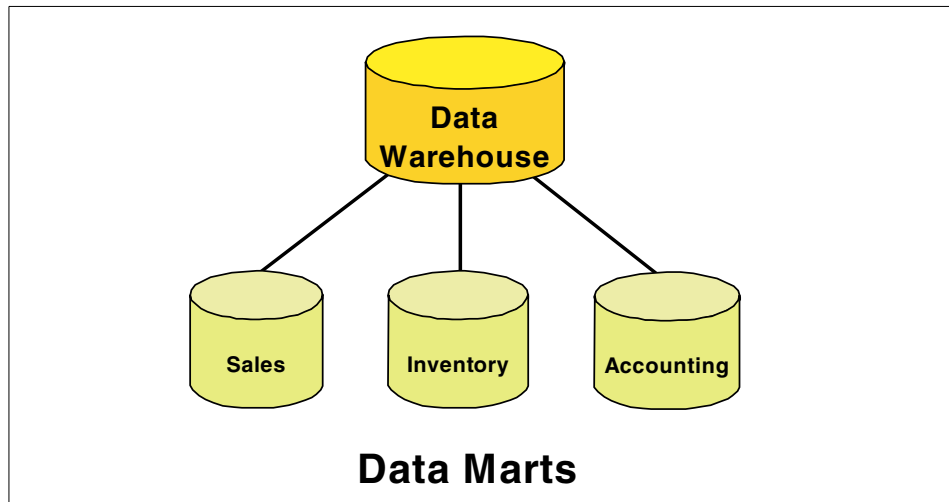
Figure 5. Data Marts representation

1.3.7 Enterprise Resource Planning (ERP)

Enterprise Resource Planning (ERP) is an evolution of MRPII (manufacturing resource planning). ERP systems are the main transactional systems used in manufacturing enterprises to automate and integrate both business and production management processes, including:

- Ordering
- Billing and shipping processing
- Integrated financial applications
- Sales and market-driven forecasting
- Materials and capacity planning
- Human resource planning
- Bills of material
- Recipe or formula management
- Integrated distribution and warehouse management

ERP has transformed the way businesses operate by eliminating barriers between functional departments and reducing duplication of efforts, but it is more than just software and hardware. ERP provides businesses with a new information technology infrastructure that encompasses hardware, software, business processes, and people.

Chapter 2. The Seascape architecture

In this chapter, we discuss IBM's Seascape architecture.

2.1 Overview

Seascape, the IBM storage enterprise architecture, is a blueprint for comprehensive storage solutions optimized for a connected world. The Seascape architecture outlines new concepts for storage by integrating leading technologies from IBM — including disk, tape, optical, powerful processors, and rich software function — to provide highly reliable, scalable, and versatile application-based storage solutions that span the range of servers from PCs to supercomputers.

Seascape solutions usher in a new era of data storage by helping your organization improve the ways it accesses, grows, manages, and secures data while protecting current storage investments. This storage architecture can help you implement a simplified, yet flexible storage infrastructure that will help you derive the most value from your information.

Highlights of Seascape are these:

- Provides comprehensive storage solutions that integrate rich software, enterprise storage management, and a full range of disk, tape, and optical storage technology
- Spans the range of servers from UNIX, Windows NT, and midrange to S/390
- Facilitates a wide array of connectivity and software functions to provide the foundation for universal data access and information sharing
- Employs "snap-in" building blocks that enable rapid integration of new technologies while protecting existing storage investments
- Provides solutions tuned to specific application needs
- Leverages industry-leading technologies that enable the development of comprehensive, cost-effective storage solutions

2.2 Seascape principles

The Seascape architecture is based on the following three major principles, described in the following sections:

- 2.2.1, "Universal data access" on page 14
- 2.2.2, "Snap-in building blocks" on page 15
- 2.2.3, "Integrated storage servers" on page 16

2.2.1 Universal data access

Seascape is focused on cutting across the boundaries and complexities of a mixed platform world to enable the sharing of storage resources and data across disparate islands of data within the enterprise (see Figure 6).

One challenge in achieving universal data access is physical connection. The Seascape storage server architecture facilitates a wide array of network and platform connectivity — including UNIX, Windows NT, midrange, and S/390 — to handle a mixed vendor environment. Another significant challenge is the requirement for a single copy of data to be used across multiple platforms for access and update. The Seascape architecture includes rich software function, which is necessary for the complex data movement and manipulation required to address this challenge.
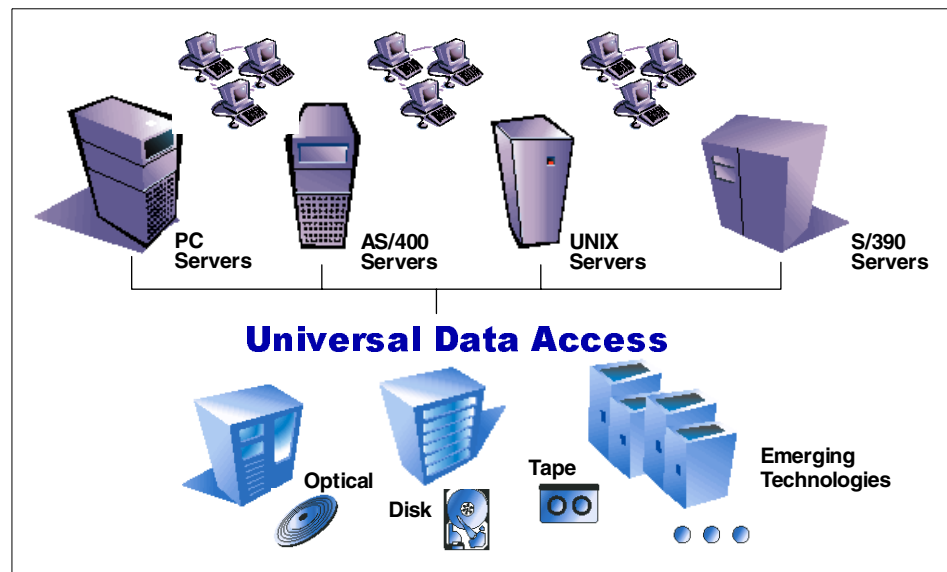
Figure 6. Universal data access

2.2.2 Snap-in building blocks

The second principle of Seascape architecture is "Snap-in" building blocks. Software and hardware building blocks are the key to keeping pace with the future, both in function and technology. This key principle enables investment protection at affordable cost. The Seascape modular architecture enables the rapid integration of new technologies (Figure 7). This modular architecture is composed of software and hardware building blocks chosen for their value to the networked model. They are the key to keeping pace with the future, enabling a new era of storage solutions that can be tuned to specific application needs.
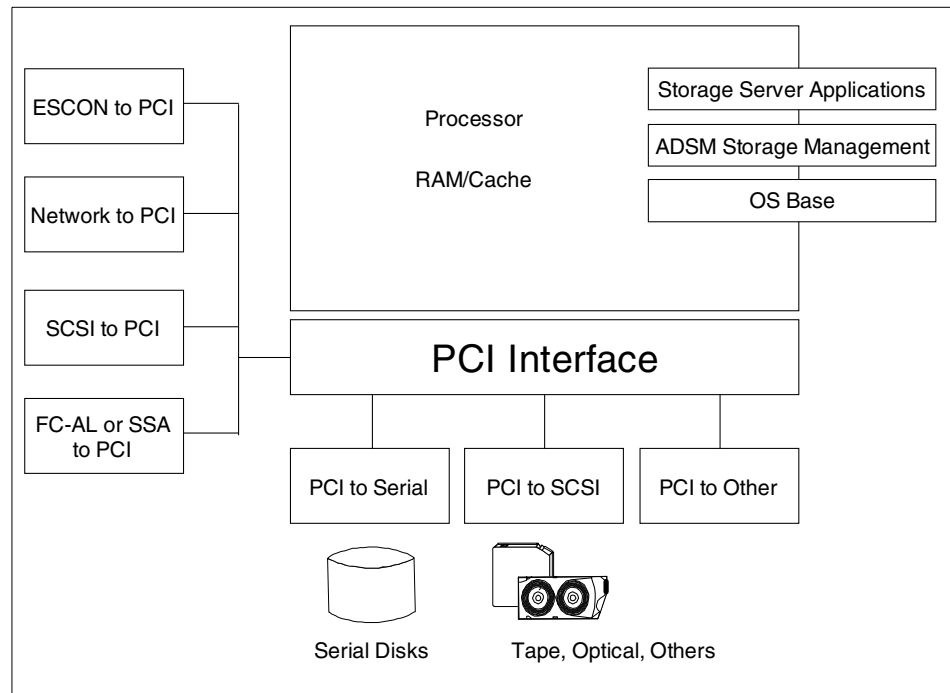
Figure 7. Seascape storage server architecture

These building blocks leverage some of the most powerful technologies available. For instance, IBM is an industry leader in storage management software, magnetic recording, and serial link technology, with innovative solutions for disk and removable media. With its significant investments in UNIX software (AIX), server architectures, microprocessors, and networking, IBM can easily incorporate new and emerging technologies into its Seascape storage solutions quickly and affordably. In fact, there is no other company

better positioned to deliver and support such a comprehensive storage architecture.

Several examples of these snap-in building blocks include:

- IBM Magstar family of high-performance tape products designed to increase capacity and provide near-on-line performance while reducing storage costs. See 8.1, "3590 High Performance Tape Subsystem" on page 153, and 8.5, "3570 Magstar MP Tape Subsystem" on page 172 for details.

- Tivoli Storage Management, IBM's powerful enterprise storage management software for distributed environments. See Chapter 12, "Tivoli Storage Manager" on page 223 for details.

- IBM 7133 Serial Disk System, a highly scalable disk system that provides high performance through serial technology. See 7.2, "7133 Serial Disk Subsystem Advanced Models D40 and T40" on page 132 for details.

The building block concept incorporates hardware and software components together in different configurations to provide storage solutions designed to fit your specific application needs.

2.2.3 Integrated storage servers

The third principle of Seascape architecture is that of the integrated storage server. To tackle the issues of complex data movement and enable new storage solutions, tuned to specific applications, Seascape architecture provides integrated storage servers. The Seascape storage servers enables a vast array of storage application possibilities. This key principle is at the heart of providing rich software function and enabling comprehensive new storage solutions. This growing library of industry software applications now opens the door for a virtually endless stream of new storage function. Storage management, backup and recovery and application accelerators to optimize database performance are just a few of the possibilities.

Seascape-compliant integrated storage servers include:

- **Network Storage Manager**, a complete solution for automated backup and recovery of corporate networks. See 10.1, "IBM 3466 Network Storage Manager" on page 205 for details.

- **Virtual Tape Server**, a hierarchical storage management system that provides virtual tape processing and tape volume stacking for users of S/390 enterprise servers. See 8.4, "Virtual Tape Server" on page 166 for details.

- **Enterprise Storage Server**, a centrally-managed disk storage solution for multiple open system and S/390 platforms. This server is also designed to ease the workload involved in server consolidation. See 7.1, "The IBM Enterprise Storage Server (ESS)" on page 121 for details.

- **7133 Serial Disk Subsystem**, a centrally-managed disk storage solution for multiple UNIX-based and Windows NT-based servers. This subsystem is also designed to allow users to ease the workload involved in server consolidation. See 7.2, "7133 Serial Disk Subsystem Advanced Models D40 and T40" on page 132 for details.

- **Modular Storage Server** (MSS) is an extremely flexible storage system designed to enable high-speed, continuous business operations for today's and tomorrow's SAN-based open environments. The MSS combines the high performance of Fibre Channel technology with scalable modular components. See 6.3 Modular Storage Server on page 112.

Chapter 3. Storage Area Network Initiative

Today's typical enterprise networks are composed by several servers with different operating systems, each of them offering specific user applications. Local data stored in one of these servers is frequently required by another application running in a different server, thus consuming servers and network resources. At this point, a new concept appears, known as Storage Area Networks (SANs).

3.1 Overview

In host-centric legacy environments, the management of storage is centralized. Storage devices are connected directly to the host and managed directly by the IT department. It is relatively straightforward and easy to manage storage in this manner.

However, the information that was centralized in a host-centric environment is now dispersed across the network and is often poorly managed and controlled. Storage devices are dispersed and connected to individual machines, capacity increases must be planned machine by machine, storage acquired for one operating system platform often cannot be used on other platforms.

Today, an IT manager's requirements are based on data availability, performance, configuration, and cost. Here is how the concept of a SAN can address these requirements:

- Facilitates universal access and sharing of resources
- Supports unpredictable, explosive information technology growth
- Provides affordable 24 x 365 availability
- Simplifies and centralizes resource management
- Improves information protection and disaster tolerance
- Enhances security and data integrity of new computing architectures
- Enables any-to-any interconnection of servers and storage systems

3.2 What is a Storage Area Network (SAN)?

One definition of SAN is a high-speed network, comparable to a LAN, that allows the establishment of direct connections between storage devices and processors (servers) centralized to the extent supported by the distance of Fibre Channel. See Figure 8.

The SAN can be viewed as an extension to the storage bus concept that enables storage devices and servers to be interconnected using similar elements as in Local Area Networks (LANs) and Wide Area Networks (WANs): routers, hubs, switches and gateways.

A SAN can be shared between servers and/or dedicated to one server. SAN interfaces can be Enterprise System Connection (ESCON), Small Computer System Interface (SCSI), Serial Storage Architecture (SSA), High Performance Parallel Interface (HIPPI), Fibre Channel (FC), or whatever new physical connectivity may emerge.
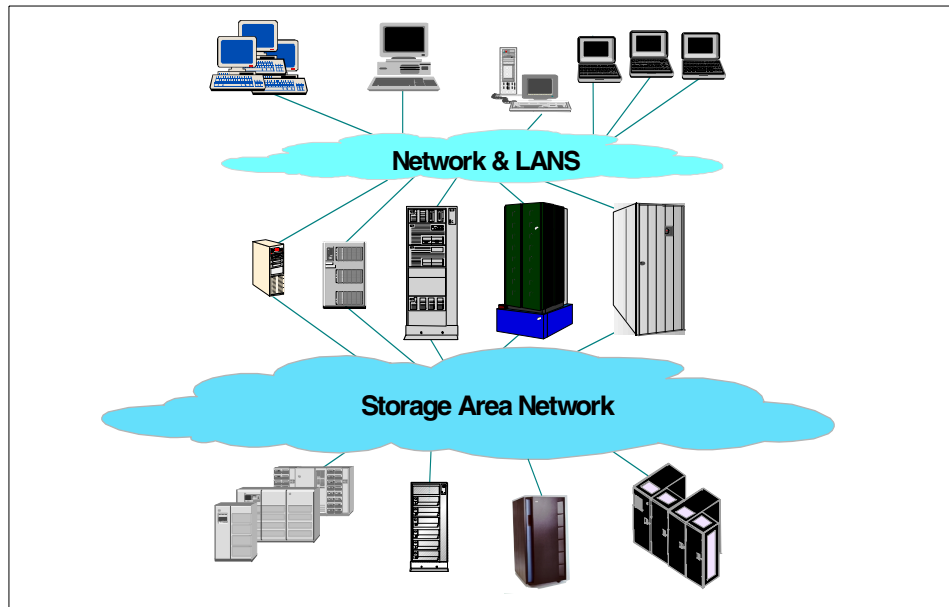
Figure 8. SAN concept

SANs create new methods of attaching storage to servers. These new methods promise great improvements in both availability and performance. Today's SANs are used to connect shared storage arrays to multiple servers, and are used by clustered servers for failover. They can interconnect mainframe disk or tape to network servers or clients, and can create parallel data paths for high bandwidth computing environments. A SAN is another network that differs from traditional networks because it is constructed from storage interfaces. Often it is referred to as the network behind the server.

A SAN can be used to bypass traditional network bottlenecks. It supports direct, high-speed data transfers between servers and storage devices in the following three ways:

- Server to storage: This is the traditional model of interaction with storage devices. The advantage is that the same storage device may be accessed serially or concurrently by multiple servers.

- Server to server: A SAN may be used for high-speed, high volume communications between servers.

- Storage to storage: This outboard data movement capability enables data to be moved without server intervention, thereby freeing up server processor cycles for other activities like application processing. Examples include a disk device backing up its data to a tape device without server intervention, or remote device mirroring across the SAN.

SANs allow applications that move data to perform better, for example by having the data sent directly from source to target device without any server intervention. SANs also enable new network architectures, where multiple hosts can access multiple storage devices connected to the same network. Using a SAN can potentially offer the following benefits:

- Improvements to application availability: Storage independent of applications and accessible through alternate data paths.

- Higher application performance: Storage processing off-loaded from servers and moved onto a separate network.

- Centralized and consolidated storage: Simpler management, scalability, flexibility, availability.

- Data transfer and vaulting to remote sites: Remote copy of data enabled for disaster protection.

- Simplified centralized management: Single image of storage media simplifies management.

Often we hear the term Network Attached Storage (NAS) or network appliance. A NAS is basically a LAN attached file server that serves files using a network protocol such as Network File System (NFS) or Simple Network Management Protocol (SNMP). It is not a SAN device.

3.3 SAN implementation

Today, the industry considers Fibre Channel (FC) as the architecture on which most SAN implementations are built. Fibre Channel is a technology standard that allows data to be transferred from one network node to another

at very high speeds. Current implementations that transfer data at rates of 100 MB/sec, 200 MB/sec, and 400 MB/sec, have already been tested.

Note: The word *Fibre* in Fibre Channel is not written using the American spelling. This is because the interconnections between nodes are not necessarily based on fiber optics, but can also be based on copper cables.

Some people refer to Fibre Channel architecture as the Fibre version of SCSI. Fibre Channel is an architecture used to carry IPI traffic, IP traffic, FICON traffic, FCP (SCSI) traffic, and possibly traffic using other protocols, all at the same level on the standard FC transport. An analogy could be ethernet, where IP, NetBIOS, and SNA all run simultaneously over a single ethernet adapter, since these are all protocols with mapping to ethernet. Similarly, there are many protocols mapped onto FC. See Figure 9.

For further information about FC, refer to 6.2.3, "Fibre Channel" on page 88.
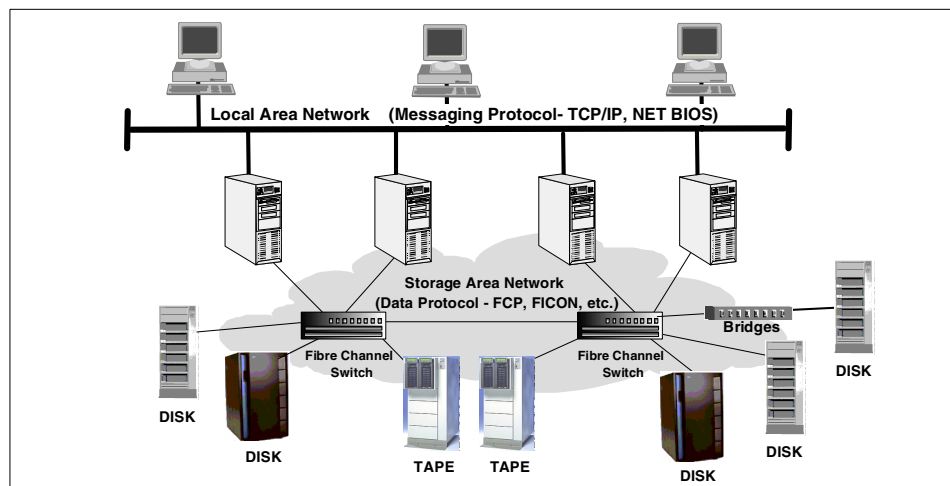
Figure 9. Fibre Channel in a SAN environment

3.4 SAN components

As mentioned earlier, the industry considers Fibre Channel as the architecture on which most SAN implementations will be built with FICON as the standard protocol for S/390, and FCP as the standard protocol for the non-S/390 systems. Based on this perception, the SAN components described below will be Fibre Channel based. See Figure 10.
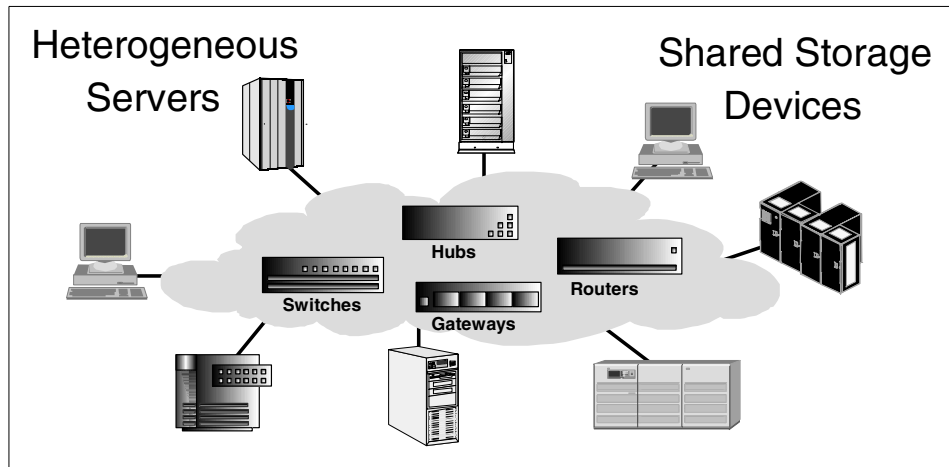
Figure 10. SAN components

3.4.1 SAN servers

The server infrastructure is the underlying reason for all SAN solutions. This infrastructure includes a mix of server platforms such as NT, UNIX (various flavors), and OS/390. Though early SAN solutions only supported homogeneous environments, SAN will evolve into a truly heterogeneous environment.

3.4.2 SAN storage

The storage infrastructure is the foundation on which information relies, and therefore must support a company's business objectives and business model. In this environment, simply deploying more and faster storage devices is not enough; a new kind of infrastructure is needed, one that provides enhanced network availability, data accessibility, and system manageability than is provided by today's infrastructure.

The SAN meets this challenge. The SAN liberates the storage device, so it is not on a particular server bus, and attaches it directly to the network. In other words, storage is externalized and functionally distributed across the organization. The SAN also enables the centralizing of storage devices and the clustering of servers, which makes for easier and less expensive administration.

3.4.3 SAN interconnects

The first element that must be considered in any SAN implementation is the connectivity of storage and server components using technologies such as Fibre Channel. Components listed below typically have been used for LAN and WAN implementations. SANs, like LANs, interconnect the storage interfaces together into many network configurations and across long distances.

3.4.3.1 Cables and connectors

Similar to parallel SCSI and traditional networking, there are different types of cables of various lengths for use in a Fibre Channel configuration. Two types of cables are supported: copper and optical (fiber). Copper cables are used for short distance (up to 30m) and can be identified by their DB9 (9-pin) connector.

Fiber cables come in two distinct types: Multi-Mode fiber (MMF) for short distances (up to 2km) and Single-Mode Fiber (SMF) for longer distances (up to 10km).

3.4.3.2 Gigabit Link Model (GLM)

Gigabit Link Models are generic Fibre Channel transceiver units that integrate the key functions necessary for installation of a Fibre Channel media interface on most systems.

3.4.3.3 Gigabit Interface Converters (GBIC)

Gigabit Interface Converters are typically used with hubs and switches, and allow both copper and fiber optics to connect up to the same hub or switch. This works well in an environment where the components to be attached may be both fiber optics and copper, depending on cost and distance requirements.

3.4.3.4 Media Interface Adapters (MIA)

Media Interface Adapters can be used to facilitate conversion of an optical interface connection to copper and vice versa. Typically, MIAs are attached to host bus adaptors, but can also be used with switches and hubs. When a hub or switch only supports copper or optical connections, MIAs can be used to convert the signal to the appropriate media type (copper or optical).

3.4.3.5 Adapters

There are devices that connect to a network, server or storage device and control the electrical protocol for communications. Adapters are also referred to as Network Interface Cards (NIC), Enterprise Systems Connection (ESCON) adapters, Host Bus Adapters (HBA), and SCSI hot bus adapters.

3.4.3.6 Extenders

Extenders are used to facilitate longer distances between nodes that exceed the theoretical maximum.

3.4.3.7 Multiplexors

Multiplexors provide for more effective utilization of high speed bandwidth resources by interleaving data from multiple sources onto a single Link. An example of this would be the Fibre CONnection Channel (FICON) Bridge, which allows up to eight separate ESCON paths to be multiplexed over a single FICON Link.

Multiplexers are also becoming increasingly efficient in terms of data compression, error correction, transmission speed and multi-drop capabilities.

3.4.3.8 Hubs

Fibre Channel hubs are used to connect up to 126 nodes into a logical loop. All connected nodes share the bandwidth of this one logical loop. Fibre Channel Arbitrated Loop (FC-AL) protocol is the most widely accepted, cost-effective alternative. See Figure 11.
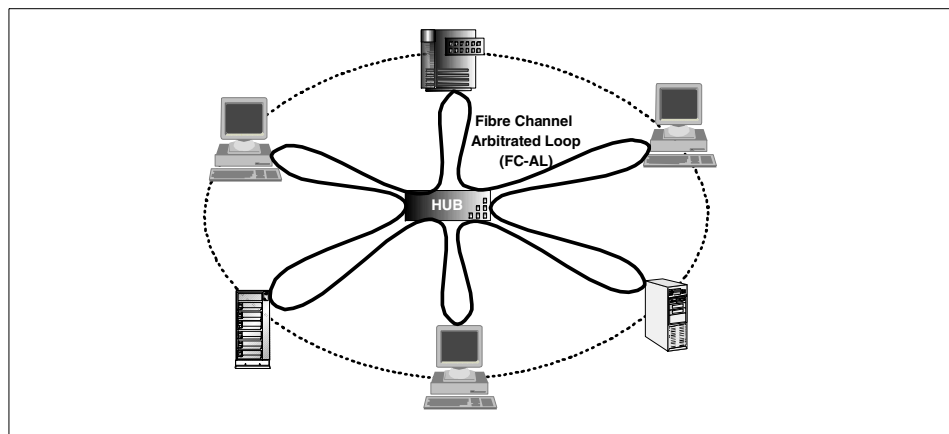
Figure 11. Fibre Channel Hub

Each port on a hub contains a Port Bypass Circuit (PBC) to automatically open and close the loop to support hot pluggability. Multiple hubs and links can be implemented to provide alternate path failover capability for high availability server environments.

3.4.3.9 Routers

Storage routing is a new technology based on the old concept of routing as it is understood by the data communications industry. Storage routers differ from network routers in that the data being routed uses storage protocols like FCP (SCSI) instead of messaging protocols such as TCP/IP.

The data path used to transfer storage data may be the same as that used for messaging traffic, but the data itself contains imbedded storage protocol information. This is similar to the concept of tunneling protocols used in the broader market. For example, a Storage Router could encapsulate SCSI information in TCP/IP packets for transmission over an intervening ethernet network.

The term routing implies data transfers over differing transmission media and addressing schemes. As a combination of communications and storage channel capabilities, Fibre Channel represents the first opportunity to apply communication techniques, such as routing, to storage traffic. Routers can be used to link LANs/SANs together locally or remotely as part of a WAN.

3.4.3.10 Bridges

Bridges facilitate communication between LAN/SAN segments and/or networks with dissimilar protocols. They are needed primarily as migration aids to the Fibre Channel environment.

An example of this would be a FICON bridge, which allows ESCON protocol to be tunneled over Fibre Channel protocol, or it can enable existing parallel SCSI devices to be accessed from Fibre Channel networks.

Figure 12 shows the IBM Data Gateway, which is actually a bridge (although it is called a gateway). This component allows up to 4 Ultra SCSI back-end interfaces and 2 front-end Fibre Channel interfaces.
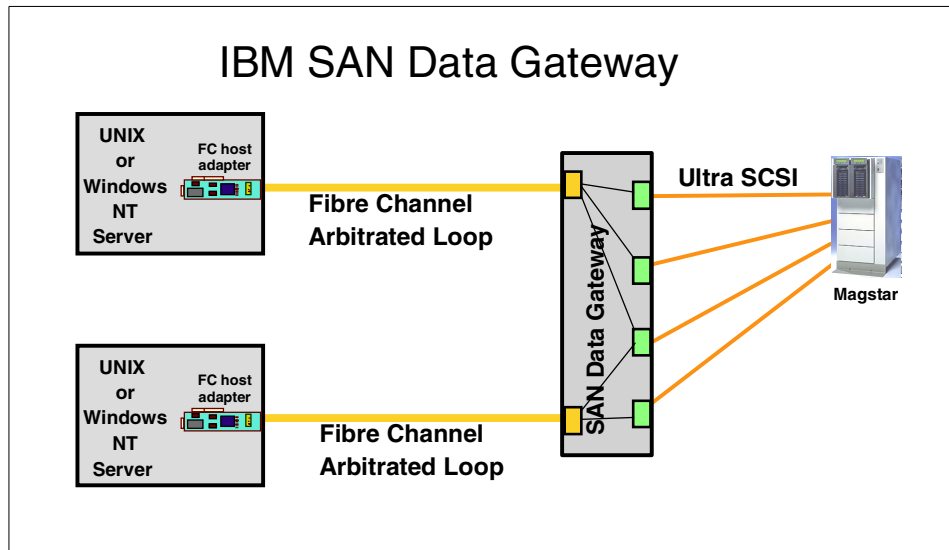
Figure 12. IBM SAN Data Gateway

3.4.3.11 Gateways

A gateway is a network station used to interconnect two or more dissimilar networks or devices, and may or may not perform protocol conversion. These boxes are typically used to provide access to a WAN from a LAN. With gateways, SANs can be extended across a WAN. See Figure 13.

Note: The IBM SAN Data Gateway is used to attach SCSI devices, like tape libraries and disk subsystems at one end, and Fibre Channel connections at the other end. This makes it a bridge, not a gateway.
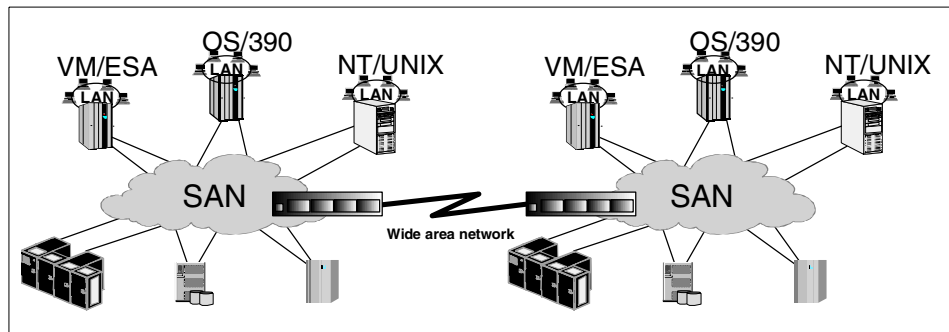
Figure 13. SAN gateway connecting remote SANs

3.4.3.12 Switches

Switches are among the highest performing devices available for interconnecting large numbers of devices, increasing bandwidth, reducing congestion, and providing aggregate throughput.

When a Fibre Channel switch is implemented in a SAN, the network is referred to as a fabric, or switched fabric. Each device is connected to a port on the switch, enabling an on-demand connection to every connected device. Various FC switch offerings support both fabric and/or loop connections. As the number of devices increases, multiple switches can be cascaded for expanded access.

3.4.3.13 Directors

Using switched technology, ESCON Directors provided the very first true homogeneous switched SAN. The ESCON Directors provide a completely separate network dedicated to move data. Data can be transferred at a rate of 17MB/sec, and dynamic connections can be established between attached ESCON control units and channels.

This is how IBM intends to announce the FICON switches:
These will be standard Fibre Channel switches with additional function to provide management from OS/390. They will provide dynamic connections between FICON channels and control units with FICON interfaces.

Since FICON switches should be standard Fibre Channel switches, non-S/390 servers and storage systems can be attached.

3.5 SAN management

In order to achieve the various benefits and features of SANs, like performance, availability, cost, scalability, and interoperability, the SAN infrastructure (switches, routers, and so on), as well as attached storage systems, must be effectively managed.

As said before, the infrastructure of SANs can be compared with that of LANs. Something similar happens with SAN management. A networking hub or switch is managed with a special software that allows, among other features, creation of several Virtual LANs (VLANs), so the communication between different LANs is restricted or allowed, providing a constant bandwidth to each port.

SAN management software provides storage devices access to the different servers allocated in the network, assigning these storage devices among several servers, sharing them between servers with the same operative system, and restricting the access to those servers which do not need it.

This software can be installed in any server or workstation connected to the SAN or by way of a TCP/IP, managing the SAN's environment with a Web browser. IBM along with Tivoli Systems, a subsidiary of IBM and a world class developer of management systems, with products like IBM StorWatch and Tivoli SAN Manager, are uniquely positioned to help you meet the complex systems and storage management challenges involved designing and implementing a storage area network (SAN).

3.6 IBM SAN strategy

Understanding that every business has unique requirements, and recognizing that investment protection and minimization of operational disruption and risk are key factors influencing an organization's purchase decision, IBM has defined a SAN strategy that:

- Uses current investments in hardware, software, and skills.

- Enables the integration of new technologies as they emerge, mature and become more affordable by deploying a building-block infrastructure.

- Applies experience and lessons learned from the mainframe arena, such as system-managed storage concepts and switched fabric management, to the open systems environment.

- Supports multi-platform, multi-vendor interoperability.

- Minimizes the risks or unproven technologies by providing "mix and match" freedom of choice to deploy what makes the most sense for each unique environment.

IBM's SAN strategy is a "whole product" strategy, an holistic approach that delivers technology, support, consulting/services and education throughout the entire SAN solution life cycle. Five basic initiatives (see Figure 14) form the cornerstone of IBM's strategy:

Hardware — To get the fullest benefit from SANs, the hardware base (servers and storage subsystems) must be SAN aware, and must provide SAN-enabled, native attachment. IBM plans to provide a full range of SAN enabled hardware, as well as "bridges" that will permit organizations to leverage existing hardware as they move toward SAN implementation.

Connectivity — Because few businesses can afford to implement a complete replacement strategy, IBM has taken investment protection and seamless migration as essential SAN design points. Therefore, IBM intends to complement its current storage/storage management, server, network and software product lines with a full set of SAN enabled products to meet specific business needs, from fabric components like gateways, hubs and switches to native Fibre Channel attached disk and tape storage devices.

Management — In the same way, a SAN management infrastructure must be able to accommodate both existing and evolving systems incorporating products from IBM and other storage vendors such as EMC and Hitachi Data systems.

Exploitation — The true value of SAN is not the specific set of technologies, but rather the software functions that use the technologies and provide business value. In its role as solution provider, IBM will exploit SAN technology to deliver a set of business solutions that address today's most pressing IT problems. Tivoli, for example, plans to deliver tape pooling, which dynamically allocates tape drives and libraries to improve efficiency, enhance productivity and provide economies of scale.

Support — SAN's promise of any-to-any connectivity and the complexity it introduces highlight the critical nature of both system integration and service in any SAN strategy. As one of the industry's leading service providers, IBM Global Services can provide the support, services and education required to support end-to-end SAN solutions.

Just as IBM is a leader in helping companies move to e-business, we will use our decades of enterprise experience and our technology leadership to become a leader in helping companies make a smooth transition to SANs. To that end, we have created a strategy that provides a complete framework filled in by IBM and non-IBM products and services. This strategy protects current investments, provides a roadmap for incorporating evolving technology, and can help minimize risk while increasing returns.
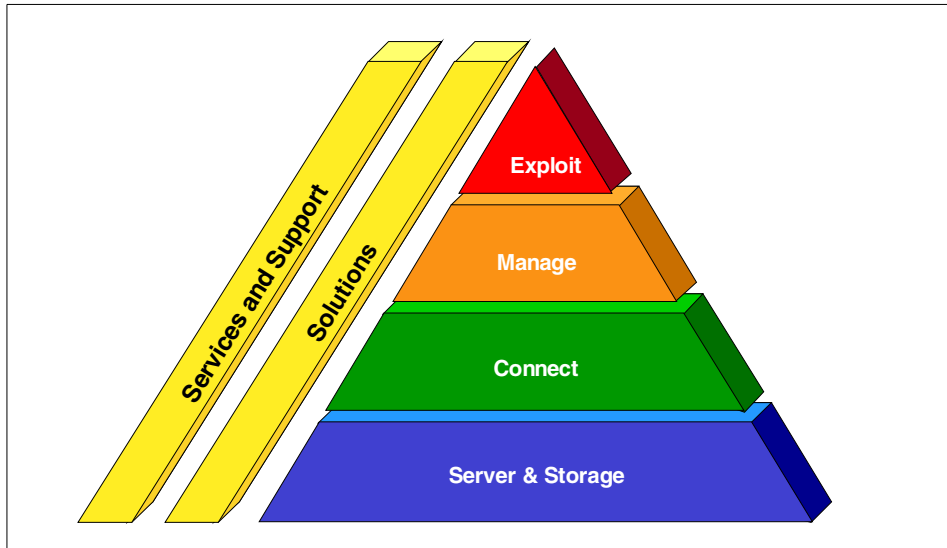
Figure 14. IBM SAN strategy

Companies that choose to begin work with IBM today to develop SAN solutions can expect to see a competitive advantage tomorrow as they create a true corporate information utility.

For further information about Storage Area Network topics, see the redbook *Introduction to SAN*, SG24-5470.

Chapter 4. Introduction to storage virtualization

In this chapter we introduce the concept of storage virtualization, and the different levels of abstraction that can be made to make storage management more efficient and transparent to the users. Virtualization is separation of the logical view of storage from its physical implementation, thus enabling physical changes to be made without changing the logical view.

For a more detailed description of storage virtualization, please refer to the redbook *Storage Networking Virtualization: What's it all about?,* SG24-6210.

4.1 The need for virtualization

Many authors these days state that storage has become a commodity over the past few years. This implies that people want to be able to just use storage without limitations or worries, to completely disregard its whereabouts and management, yet always be sure of its abundance and availability. At the same time that storage cost has been steadily going down, because of technology evolution and new ways of connecting storage devices have been evaluated and implemented, the volume of data storage required in daily life and business has exploded. The capacity is growing by 60% and hardware cost is decreasing by 30% per year, but availability requirements are approaching 100%. Three hundred million Internet users are driving two petabytes of data traffic per month. Users are mobile, access patterns unpredictable, the content of data becomes more and more interactive, so a new level of bandwidth and load balancing are required.

Managing these high volumes of critical data has become impossible as a manual task, and automatization of the processes has been one of the major efforts in the industry in the last decades. Especially in the Open Systems environment (UNIX and NT), this has introduced new policies which used to be the discipline in the mainframe glasshouses before.

Sharing of data between heterogeneous platforms also has become a prime concern, since the communication networks cannot cope with the massive replications, downloads and copying that are required in today's industry.

The development of Storage Area Networks (SANs) has passed its experimental phase and SAN has become a mature technology that allows businesses to implement storage pooling and sharing today. SANs are high speed switched networks designed to allow multiple computers to have shared access to many storage devices.

Yet, implementing a SAN does not in itself solve the basic problems of meeting the requirements of the ever growing explosion of data. Sophisticated software architectures will be needed to handle them and take advantage of the hardware.

Similar to the introduction of virtual memory in computer science that has solved all the problems of managing a limited resource in application programs, the virtualization of storage can give users a more intuitive usage of storage, while it is being taken care of transparently by the software that manages the storage network.

4.2 A definition of virtualization

A few years ago, "virtual storage" was a synonym used for the virtual memory techniques used in operating systems. Consultants and industry have begun to use the term storage virtualization to describe the abstraction from physical volumes of data storage to a logical level recently, and the new buzzword has been rapidly accepted by the IT profession. Even though it is a word that does not exist in any English dictionary, we will continue to use it since the industry has embraced the term. In proper English it could be defined as vulgarization of storage, meaning wide diffusion and popularization of storage.

This abstraction can be made on several levels of the components of storage networks and is not limited to the mere disk subsystem. Virtualization separates the representation of storage to the operating system and its users from the actual physical components. This has been represented in the mainframe environment for many years by such methodologies as system-managed storage, and products like DFSMS.

4.2.1 Levels of virtualization

We will define the different levels at which virtualization can be achieved in a storage network, as illustrated in Figure 15.
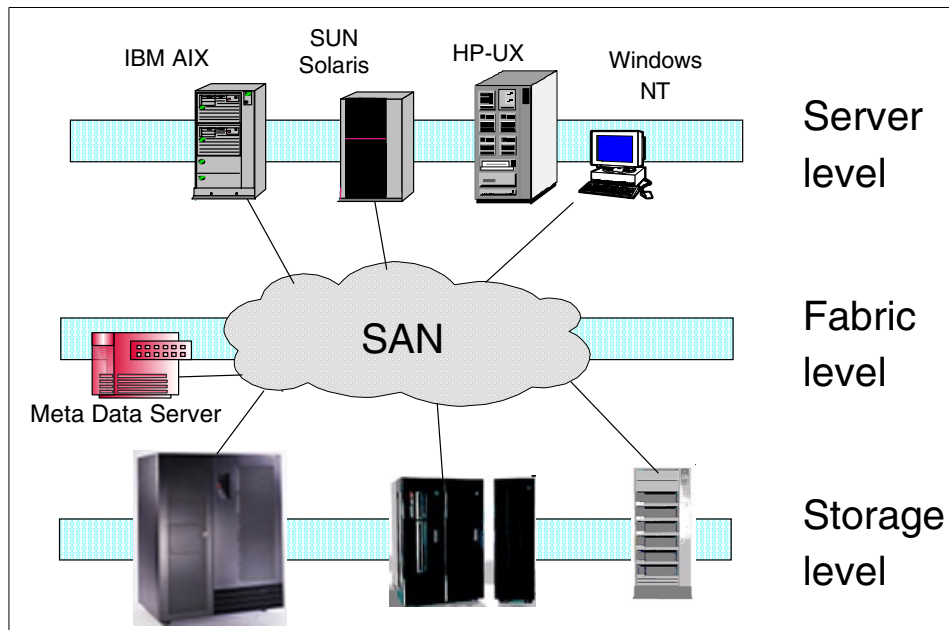
Figure 15. Levels at which virtualization can be applied

Server level

Abstraction at the server level is by means of the volume management of the operating systems on the servers. Some operating systems, like OS/390 with DFSMS, have a high level of abstraction, whereas others still present to the user volumes that are directly related to physical disks. At first sight, increasing the level of abstraction on the server seems well suited for environments without storage networks, but this can be vitally important in storage networks too, if we look at the security and sharing issues.

Fabric level

On the SAN Fabric level, virtualization can enable the independence of storage pools from heterogeneous servers. A single management interface can be implemented to manage disparate storage systems without impacting the servers. Several approaches are possible, and we will go into more detail on this.

Storage level

At the storage device level, several virtualization techniques have already been tried, ranging from striping and redundant array of independent disks (RAID) to sophisticated volume management in the Enterprise Storage Server (ESS) or Log Structured Arrays (LSAs), such as the RAMAC Virtual

Array (RVA). The Virtual Tape Server (VTS) is another example of virtualization at the device level. Future developments can implement virtualization at this level to enhance functions like cloning or snapshots and allow storage usage beyond the physically installed capacity.

We mainly address storage virtualization on SANs in this book, but the Network Attached Storage (NAS) products can play a role in this arena as well. They provide easy to manage storage attached directly on the communications network, and basically act like more or less sophisticated file servers. Actually, they are virtual disk servers on a separate network, the Internet/intranet or the LAN/WAN.

File system level
The highest level of abstraction, to the file system level, will result in the greatest benefit, since what we want to share, allocate and protect are not volumes, but data sets or files. Therefore, a model that can make abstraction to the file level or even to a higher granularity, like to a block, will yield the best results. This may seem to go against the trend to consolidate storage at the device level, but it introduces the global data space across heterogeneous networked storage devices. Both SAN and NAS storage products can be part of this space.

Scalable, managed distributed file systems must be developed, without performance penalty from client/server interactions and with integrated security, data integrity and continuous availability.

4.3 The purpose of virtualization

Storage network virtualization addresses the increasing complexity of managing storage and will reduce the associated costs dramatically.

Its main purpose is the full exploitation of the benefits promised by a SAN, like data sharing, higher availability, disaster tolerance, improved performance, consolidation of resources, and many more, which do not automatically result from the SAN hardware components and which were so far only partially addressed by the available software that are often limited to a subset of supported platforms.

The objectives are:

- To remove the management of the storage devices from the server platforms
- To allow management by high level attributes rather than blunt physical characteristics

- Allocation and management in accordance to the Quality of Service (QoS) associated with the data

- To effectively pool or group heterogeneous storage devices as a single resource

- To adapt the size of the pools to varying capacity requirements

- To migrate data within the pools to match its profile in performance, availability, frequency of access, and so on, as pools expand or retract

- To allow for load balancing and avoid introducing hot spots in the data path

- To efficiently assure the availability of data by using mirrors or snapshots to speed up recovery and backup

- To present one single management interface to the administrator(s) which is easy to learn and use

- To be completely transparent to various applications and heterogeneous file systems

- To provide continuous availability of the virtualization system by failover

- To coexist with non virtualized storage on the same SAN and evolve with the introduction of new technology

By realizing these objectives, storage virtualization can reduce costs associated with storage management:

- Better usage of the available storage capacity in a true open SAN environment

- Better usage and development of the skills available to manage the SAN allowing administrators to manage larger amounts of storage

- Reduced human error by automatization

- Enforcement of rules concerning storage usage

Translated to business value, this signifies:

- Ability to anticipate trends

- Get to market faster

- Take advantage of new opportunities

- Minimize business risks and exposures

- Gain competitive advantage

- Exploit technology change

- Handle growth

- Reduce cost and improve profit

4.4 Virtualization models

Symmetric and asymmetric models can be drawn for storage virtualization, as illustrated in Figure 16.

File systems can be distributed in either way. When you look at Figure 16, you may instinctively conclude that asymmetrical is better than symmetrical. Nothing is less true. Both have advantages which we will illustrate in this section.
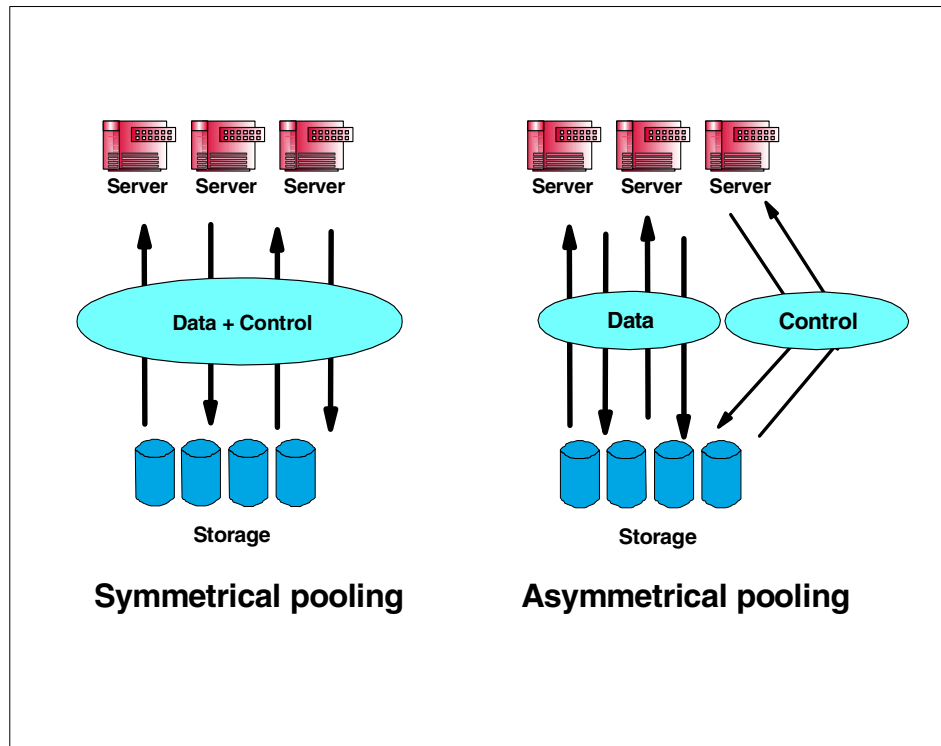
Figure 16. Symmetrical compared to asymmetrical pooling

4.4.1 Terminology

Before we continue with this discussion, let us clarify some terms. Since we are used to the terms *client* and *server* in application development, things can

get very confusing if we use them in another context. The notions of client, server, controller, agent and so on are usually associated with a certain hierarchy of service execution. Depending on the paper you read, you may see different terms or acronyms describing the same things. Some people call it a *metadata server*, while others mean a similar thing when they write *storage domain server* or *storage domain controller*.

In this book we use the following terms:

Host   Any server that is using storage and runs applications for its clients.

Agent   A hardware or software component installed on a host to virtualize its view of storage or to execute a specific storage management function for that particular host.

Server   A component of the storage network that provides a service to its clients. This can be a host server or a metadata server.

Client   Something requesting a service from a server. This can be a client to an application on a host or a host when requesting service from a virtualization server.

Controller A hardware or software component that controls access and usage of storage resources it manages.

Now to save you from total confusion, let us agree that client and server are interchangeable terms and that the relationship is defined by the interaction at a particular time. We will use illustrations to clarify these interactions and qualify the terms, where appropriate.

4.4.2 Symmetrical virtualization

When we implement a virtual storage network in a symmetrical way, both data and control flow over the same path. Levels of abstraction exist in the data path, and storage can be pooled under the control of a domain manager. Logical to physical mapping is done by the domain manager or metadata controller. Some authors will tell you that the only advantage of this approach is that it is simple, but the disadvantages lie in the single points of failure or performance bottlenecks that it can introduce. Also, it may be more difficult to move data across pools if it is controlled by different managers.

On the contrary, locking and data sharing integrity can be implemented in a simple fashion, using zoning or masking techniques that are available today. Symmetrical virtualization allows you to build a kind of fire wall in the storage network. Also the metadata controllers or domain managers can implement security, caching, snapshots, remote copy and many other functions, since

they directly control access to the storage and do not only grant access to it. Figure 17 illustrates these characteristics of symmetrical virtualization.
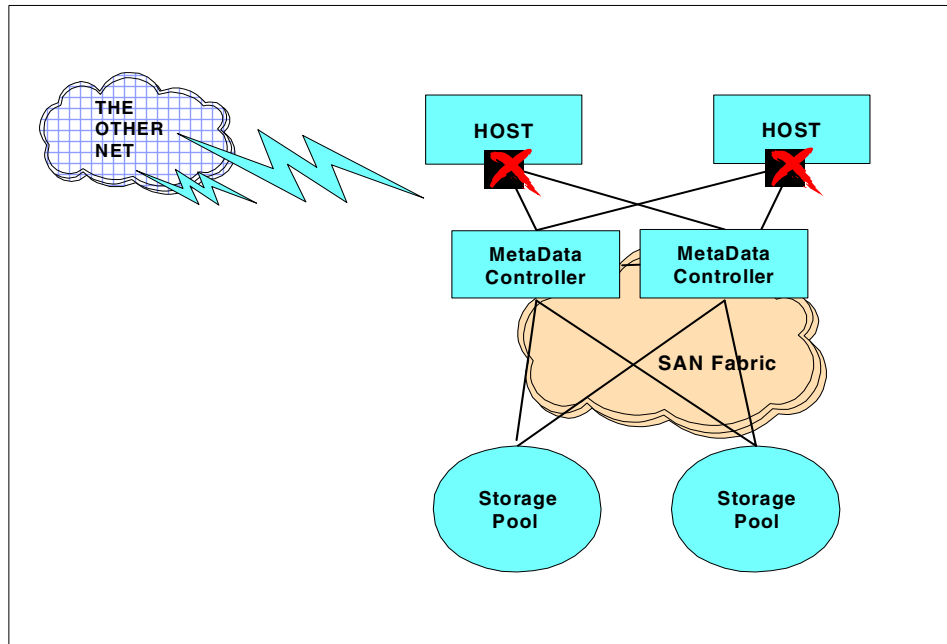
Figure 17. Symmetrical virtualization

When it comes to performance, it is all a matter of scalability. Who said there should be only one single metadata controller (MDC) in the SAN clustered with a failover? For all we know, there could be any number of cooperating domain controllers, in an n-way cluster. Adding processing power, cache memory and adapter bandwidth can be scaled to achieve the performance level required. And, this cache and processing power can be used to implement functions like snapshots and caching on devices like "just a bunch of disks" (JBODs) or RAID arrays that do not have these functions.

As the cross marks in Figure 17 indicate, if the host bus adapter (HBA) is sufficiently fenced, a firewall against unsolicited access is easily built at this level.

4.4.3 Asymmetrical virtualization

In asymmetrical virtual storage networks, the data flow is separated from the control flow. This is achieved by storing data only on the storage devices, and moving all mapping and locking tables to a separate server that contains the metadata of the files. This approach is certainly more complex to implement,

but has its strong points as well. Figure 18 illustrates asymmetrical virtualization in a different way.
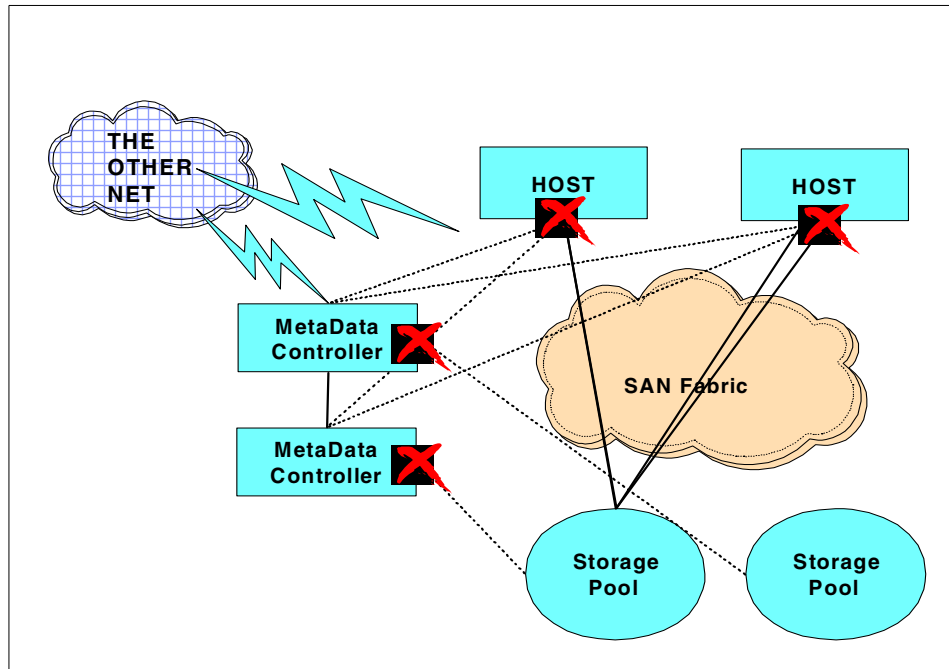
Figure 18. Asymmetrical virtualization

Separating the flow of control and data allows the IO to use the full bandwidth of the SAN, while control could go over a separate network or over circuits in the SAN that are isolated for this purpose. For many operations, the metadata controller does not even intervene. Once a client has obtained access to a file, all IO will go directly over the SAN to the storage devices. However, note the increase of security exposures this introduces — a separate control network should also be firewalled.

4.4.4 Combinations and variations on these themes

We can take this one step further and separate the SAN into two separate networks, an upper one to connect the hosts to the domain controllers and a lower one to connect the domain controllers to the storage devices, as shown in Figure 19.
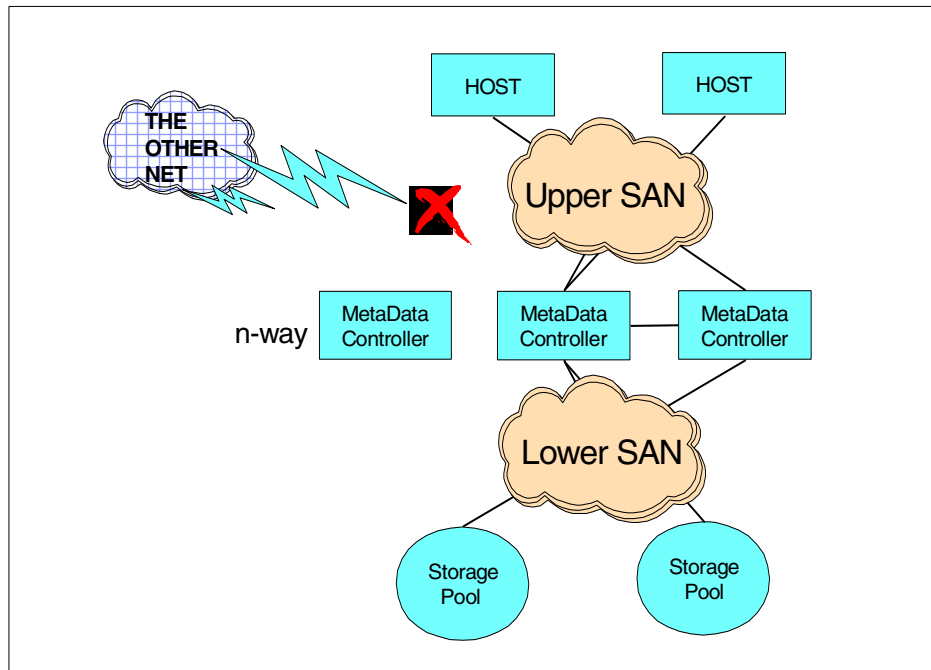
Figure 19. Two-layered storage network

In this way, we can scale performance, bandwidth, processing power, caching in an n-way cluster of domain controllers, who do virtualization both at the fabric and at the storage level, and the security firewall can be set up in a symmetrical way.

Taking another view, why would we not combine asymmetrical and symmetrical virtualization? Not everyone uses remote copy, FlashCopy, snapshots, or caching on every piece of storage equipment they bought. After all, virtualization is trying to set up different QoS levels on heterogeneous storage devices. So we can draw Figure 20 with a combination of symmetrical virtualization through a domain server that has high performance, high availability functions for one storage group, and through an asymmetrical virtualization engine for basic QoS management. If functionality is needed, both could communicate extent lists to invoke each others functions for the files they manage.
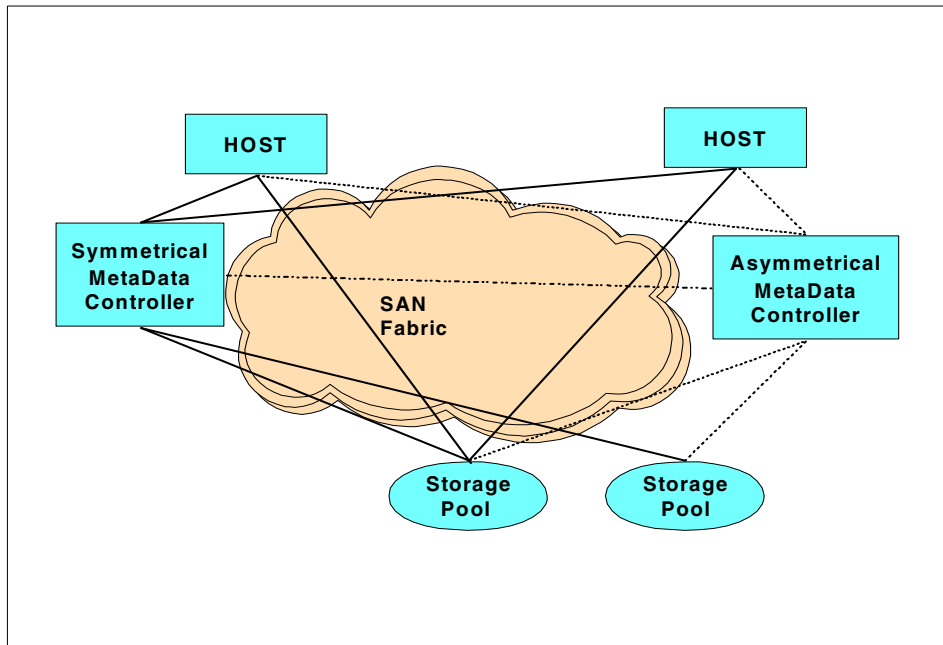
Figure 20. Combining both approaches

We will come back to these possible combinations and variations after an overview of the technology that will become available soon.

4.4.5 IBM SAN network software model

Whether symmetrical or asymmetrical virtualization is chosen, software is the prime element. IBM/Tivoli's storage network software model is illustrated in Figure 21 as a five layer model encompassing all aspects of systems management.
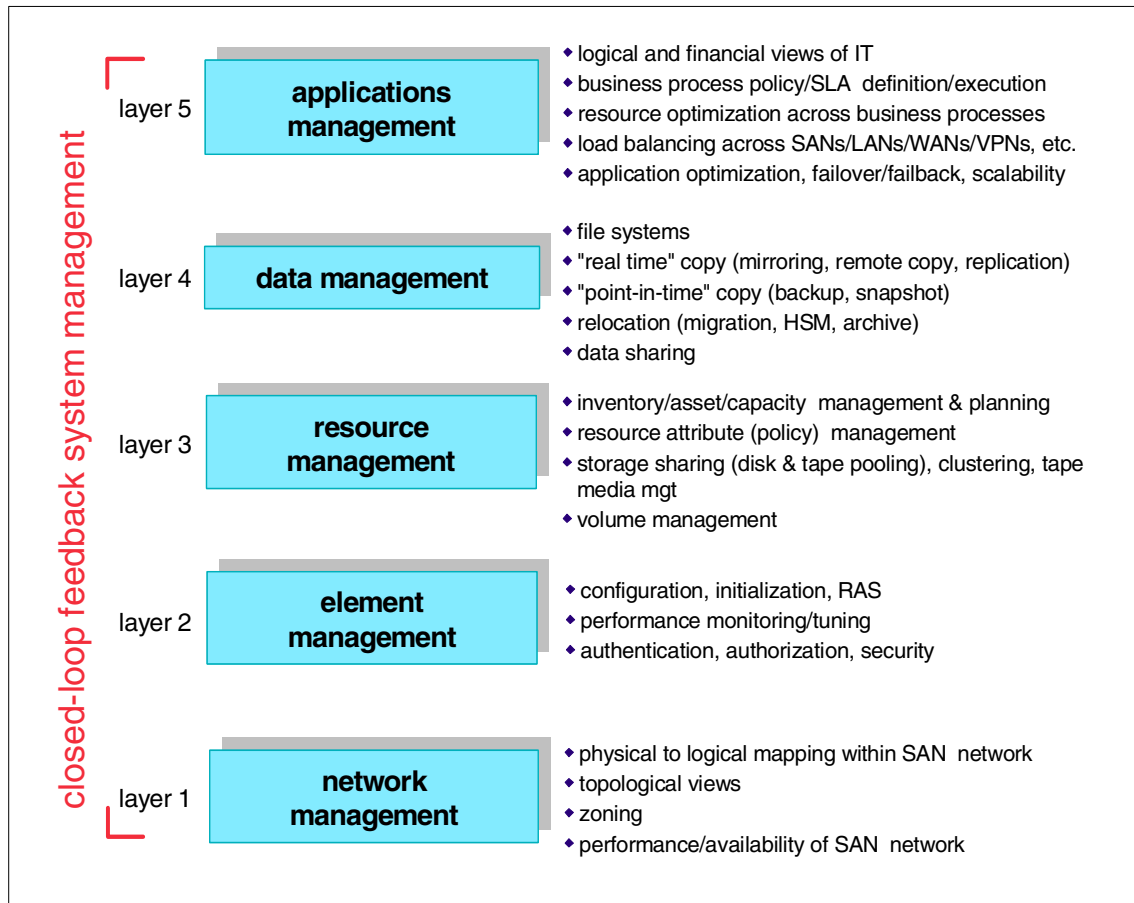
Figure 21. IBM/Tivoli SAN network software model

These functional layers allow us to implement a storage software system structure that isolates the application from the underlying storage hardware, and it can be used for file type applications as well as for database applications, as is illustrated in Figure 22.
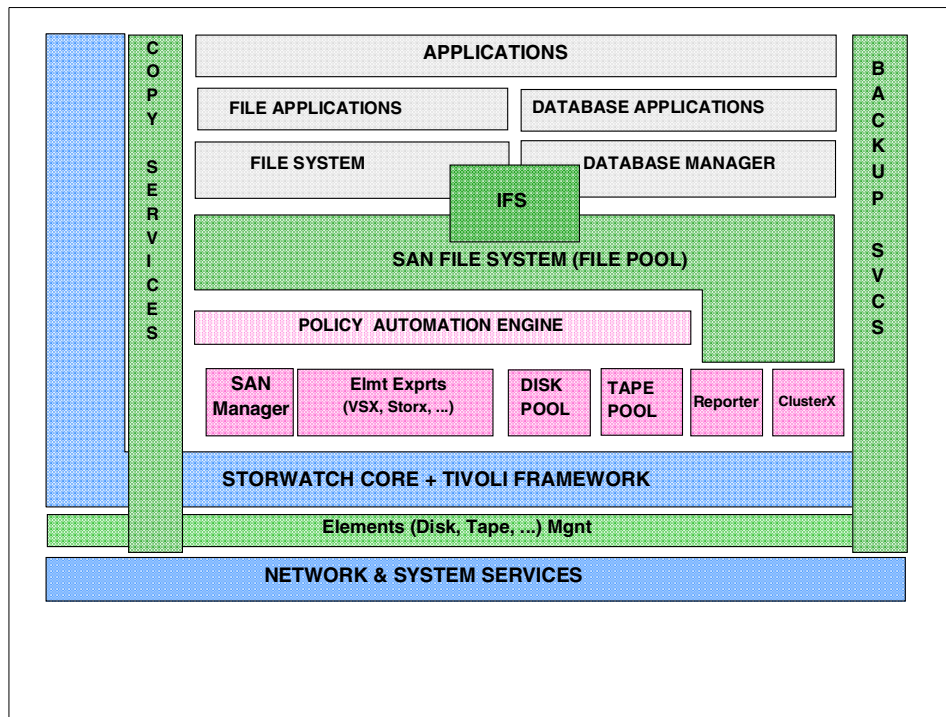
Figure 22. SAN software system structure

Note that in this model, both file system and raw device level access should be possible. This means virtualization to a higher level of granularity for some applications, whereas others continue to use volumes, even if they are managed by the virtualization software. The applications and the operating systems on the host servers are isolated from direct volume management by an Installable File System (IFS) which can either supplement or completely replace the volume management of the host server (UNIX, Linux, NT).

4.5 Current status

At this moment, several virtualization systems have been designed at various levels (server, fabric and storage). Most have been described in white papers and research reports, and some are close to being turned into new products or new releases.

In this introduction we want to briefly cover these as illustrations of different levels of abstraction that are under development today.

Under the IBM-Compaq agreement, IBM has announced that it will endorse the VersaStor technology originally developed by Compaq.

To position the different products, we have tried to put some of these in the illustration in Figure 23.
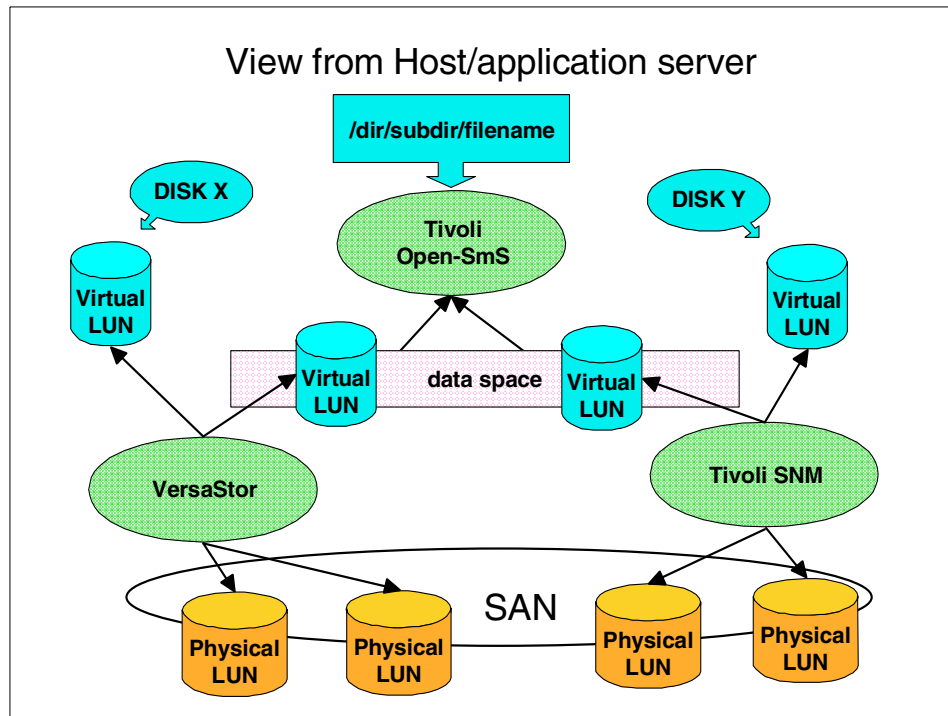
Figure 23. Positioning of architectures and products

We should discern between virtualization models that come in at the highest level, the file system, and virtualization appliances that only present volumes to the operating system, which can then use these in its volume management.

Tivoli Open System-managed Storage (SmS)
Tivoli Open SmS is a new storage management model that implements virtualization in a heterogeneous SAN environment, independent of hardware. It comes in at the highest level and sets up one single distributed file system that can be shared and used by all the servers that plug in to the storage utility. Tivoli Open SmS sets up storage groups spanning pools that have been defined across storage devices on the SAN.

Because it is hardware independent, it could use any storage from JBODs, RAID disks to ESS or MSS-like subsystems, and even virtual volumes that are managed by another virtualization technology, such as VersaStor, or a SAN disk manager, such as Tivoli Storage Network Manager.

The hosts see storage groups that can be segmented according to policies for QoS. One example could be: high, medium and low performance or availability storage groups. Or in another installation: project A, project B, project C could each have their allocations in subsets of the storage groups. Who actually can access a particular file in the shared groups is a matter of security, and Tivoli Open SmS manages this via a POSIX compliant access control. See Figure 24.
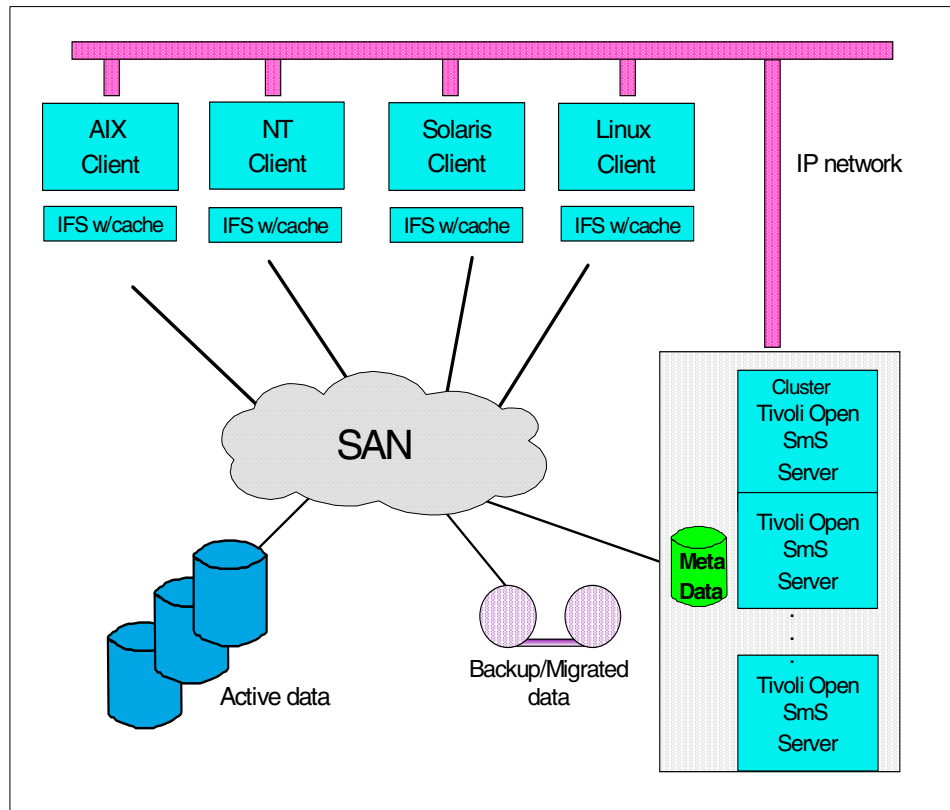
Figure 24. Tivoli Open SmS

VersaStor

VersaStor is a storage virtualization technology that uses both software and specific hardware components and virtualizes volumes at the SAN fabric level.

Between the hosts at the top and the storage devices at the bottom, it creates a set of logical volumes out of the storage volumes it finds at the bottom, and manages these via an asymmetric control unit.

VersaStor separates the traditional control unit function from the storage device and sets it on a separate control unit on the SAN, therefore allowing virtualization across theoretically any kind of disks.

It hands the volumes it defines over the SAN to the attached host servers, who then manage it as a volume in their volume management, but cannot share it with other host servers, as such. VersaStor does not handle SAN topology or resource management. It does implement virtual volumes and it isolates the host servers not only by the software, but also by using a specific VersaStor enabled host bus adapter.

Tivoli Storage Network Manager

This SAN software suite encompasses the products previously called:

- SAN Manager
- Disk Manager
- Storage Automation

Tivoli Network Storage Manager provides the vital SAN management that virtualization systems rely on. It leverages Tivoli's Common Services and uses presentation services based on a Netview interface. The administrators can launch the different element managers from the common console to manage SAN fabric elements, such as switches or hubs and storage elements like disks, tapes, optical and robotics.

SAN topology discovery and monitoring are provided both in-band and out-band to assist in SAN infrastructure maintenance and improve availability for the applications.

Disk resources can be identified and allocated across the SAN in a secure and simple way on heterogeneous disk subsystems.

Information on the elements is stored in a database which can be used in data mining by the Tivoli Storage Resource Manager for trend analysis.

4.6 Network Attached Storage (NAS)

NAS devices do file virtualization on the communications network. As illustrated in Figure 25, they come in two flavors:

- NAS Appliances which contain their own disk storage, protected by RAID
- NAS Gateways to other disk subsystems, which can be on a SAN
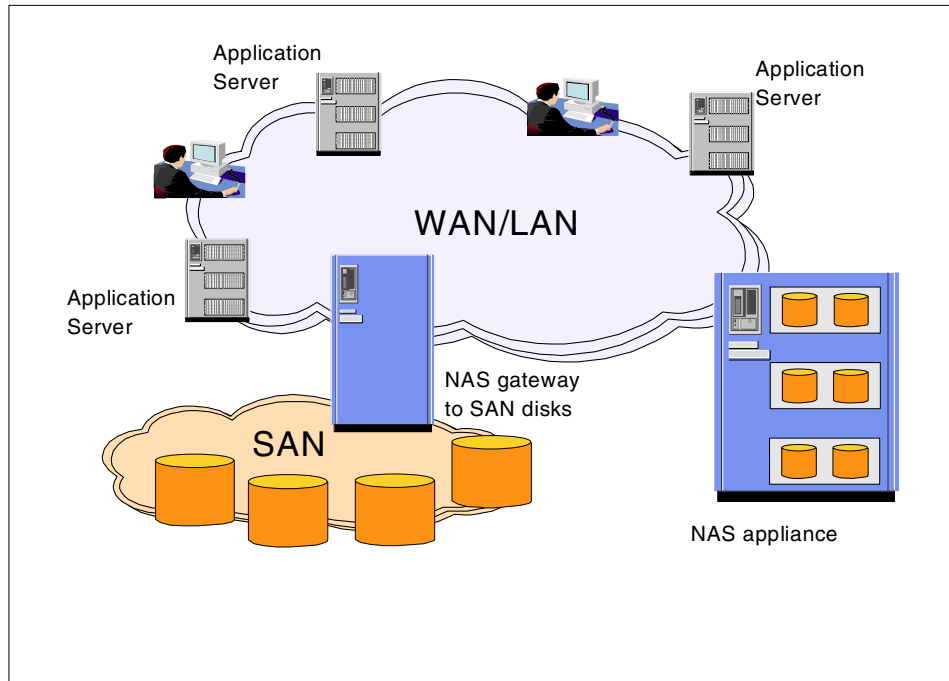
Figure 25. NAS storage devices

A NAS device is a virtual file server, and supports NFS-like protocols. To position NAS versus SAN devices, we consider several arguments, as listed in Table 1.

Table 1. NAS, SAN, and direct attached storage

Direct attach	SAN	NAS
Low configuration flexibility	Highly scalable	Low scalability
High administration costs	High IT skills needed	Lower IT skills needed
No data sharing	Limited data sharing unless virtualized	File sharing

Direct attach	SAN	NAS
High performance	High performance	Limited performance
Limited scalability	Scales UP	Scales DOWN
Inexpensive to install	Lengthy, expensive initial deployment	Quick, inexpensive initial deployment
No other connectivity	Limited reach	WAN reach

NAS and SAN use different protocols, but they will probably converge one day. The gateways are a first step. For many installations, NAS will be a first step in storage networking.

You may not want to deploy a SAN in each location, so NAS could be a complementary solution for remote locations, even if that creates new islands of information. Most virtualization products promise a future integration of NAS.

4.7 iSCSI - SCSI over TCP/IP networks

Since the bandwidths of both SANs and Ethernet networks have gone over the gigabit limit, many people are beginning to consider using cheaper Ethernet interfaces rather than the expensive Fibre Channel interfaces. Another argument is that every server will need an Ethernet (or other TCP/IP) adapter anyway, so why not use it for other purposes? And, skills to manage an IP network are more readily found than skills to design and implement a SAN.

This could provide a new way to bridge distances by using the communications network for storage networking.

However, communications networking is based on design considerations different from those of storage networking. Is it possible to merge the two and yet provide the performance of a specialized storage protocol? The challenges are many. The TCP/IP protocol is software-based and geared towards unsolicited packets. On the other hand, storage protocols are hardware-based and are based on solicited packets. Storage wants acknowledgment of all commands or data sent — communications will simply resend if no answer comes back. Data integrity must be handled with care.

A storage networking protocol would have to interface with TCP/IP without change (it's a standard) and yet achieve high performance. Storage block IO could be slowed down by sending it over a different type of network. And, is it

really a good idea to mix traffic on a single network? One of the first aims of SAN is to free up LAN bandwidth, remember?

On the other hand, much of the IP functionality already has moved to the adapter level, which frees up processor time on the servers. And the IP network could be split in separate communications and storage IP networks, or different priorities and class of service for both types of traffic can be used within a single network.

Figure 26 illustrates the possibilities of iSCSI, by itself, or integrated with SAN and NAS products.
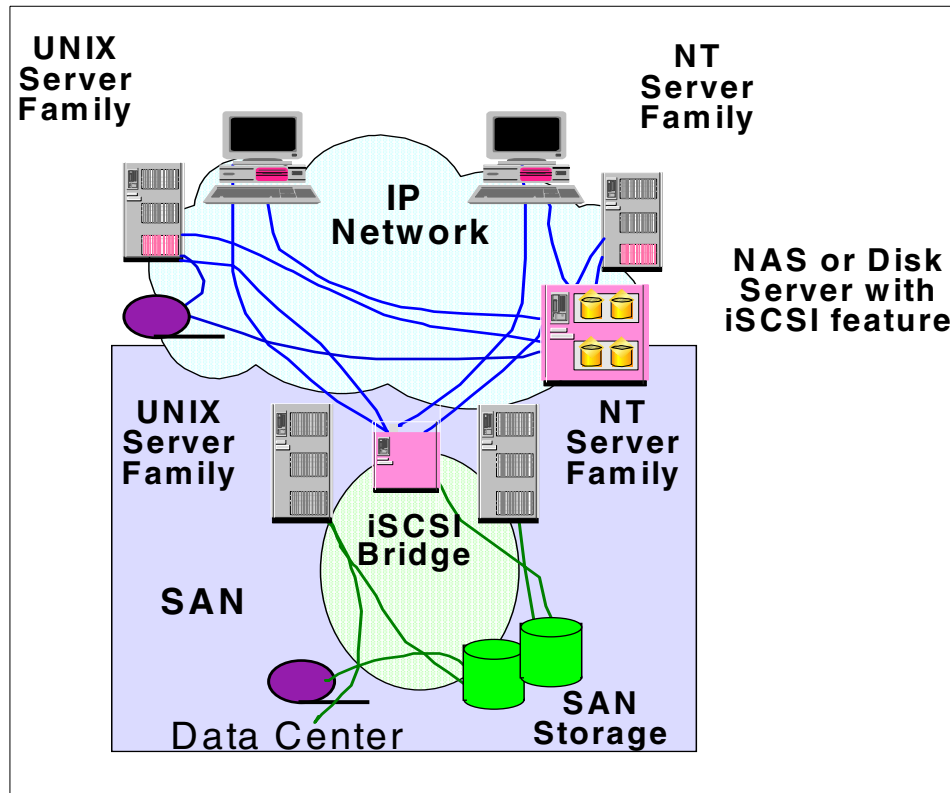
Figure 26. iSCSI in storage networking

IBM is investing in research on iSCSI to address these challenges and make it a reality, and to develop products (iSCSI interfaces on storage devices, device drivers) and management software for this new environment.

4.8 The future

When deciding on a virtualization approach, you should consider the ability of the model to evolve with new technology and the associated appliances and elements, whether it's hardware or software. If you choose for a model that will lock you into a platform limitation or even a single vendor, then going to another virtualization approach in the future may prove difficult.

If you choose a model that does not put limitations on the hardware and can use other software as an engine for certain functionality, like specific device configuration management, then that will set you on a path that allows you to integrate future technologies.

Like in basic SAN component decisions, openness and standardization are key factors. Ultimately what we really want is a universal file system.

4.9 Benefits

In this section we point out some of the benefits of virtualization.

Reducing cost
Over the last several years, it has been widely documented and understood that the cost of managing storage will easily outweigh the cost of purchasing the storage hardware.

Saving time
As corporate data assets continue to grow equating to an ever-increasing necessity for data storage resources, the time demands for planning, upkeep and maintenance by the IT staff are juggled and stretched to the breaking point. One of the most time-consuming operations in storage is managing the necessary addition of new storage hardware, the transfer of information from old to new, and the removal of old or obsolete hardware.

Managing your data storage resources
Most businesses are concerned with the efficient utilization of their data storage resources. The entire Storage Resource Management industry exists to enable IT organizations with answering usage questions.

How much storage do I own? How much has been assigned to hosts? Of that, how much has been formatted by a System Administrator? After formatting, how much usable space do I have left? How much of the formatted space has been given to a specific file system. How much of that space are the individual file systems actually using? Where am I running low? Where am I over allocated? The questions that IT organizations are asking point out that

in open system environments, the basic model of a host system owning storage leads to fragmentation, inefficient usage and difficulty in tracking. The problem is that even when the questions are answered, technology is not readily available to solve the problems.

In sharp contrast, the complete virtualization of disk storage into QoS pools of capacity provided by System-managed Storage eliminates the kind of fragmentation and inefficient usage found in open system environments. All participating host systems share the capacity. Capacity is only used as new data assets are created. Out-of-space conditions are virtually eliminated through the use of intelligent overflow capacity and the implementation of capacity-on-demand.

System-managed Storage utilizes data storage resources much more efficiently on OS/390 when compared to current open system volume managers and file systems. Storage virtualization changes this by introducing to open system environments the validated data storage techniques that have been available for years in OS/390 environments.

4.10 Deployment

In this section we will highlight some considerations related to the implementation of virtualization.

4.10.1 Time frame

Clearly, many virtualization technologies are in the phase of moving from research into development. True virtualization products can be expected to release a first version in the year 2001.

This should not stop you from moving in the direction of virtualization, since the most important job is to clearly define your company's policies on data management, and bring law and order in your SAN.

4.10.2 Configuration considerations

Let us briefly discuss some possible implications.

4.10.2.1 HBAs
A virtualization system may depend on specific hardware on the servers that plug in to it. This is true for VersaStor, which will rely on specific HBAs to enable it. Other systems, like Tivoli Open SmS, can be implemented without any specific hardware requirements.

4.10.2.2 File systems

The ultimate purpose of virtualization is to provide true universal data sharing. This may not be possible today across all existing platforms. The file systems in heterogeneous systems differ. Even when servers address NAS, which acts as a network file server, the protocols used can vary: Windows NT uses Common Internet File System (CIFS) and UNIX uses Network File System (NFS). The access control and locking that the metadata server needs to provide is based on Access Control Lists (ACLs), which are different between NT and UNIX.

Using an Installable File System (IFS) to mask these differences in operating systems, is a very open approach to solving this.

4.10.2.3 Availability

One of the major concerns is to implement a system that truly enhances the availability of data. At a workgroup or department level, this can mean 99.9% availability (meaning an average downtime of 8.45 hours per year). At the enterprise scale, we need systems that have a 99.999% availability with less than five minutes planned or unplanned outage. Designing the storage network for either situation will have different implications and cost.

Failover

Any new server or appliance that is introduced in the storage network to implement virtualization, manages metadata, pools or volumes should never become a single point of failure. Dual active servers clustering or other rapid failover is a prime requirement.

Scalability

Adding storage capacity or reconfiguring it must never imply an outage of the storage already in use. The virtualization system should be designed for easy growth, introduction of new technology and easy migration.

Dynamic growth and reduction

As the requirements change, the system should adapt itself, by moving data transparently from one device or pool to another. Storage groups are able to expand and shrink as devices are added or removed. Cloning and snapshots can be used to assist in data movement across the pools. If these functions are not available on some of the storage systems in the configuration, then the virtualization software can initiate and execute such copies by itself. Isolating the servers and applications from the management of the storage allows the virtualization system to change allocations and adapt the metadata without impact on the users.

Data movement
The virtualization system should be able to move data across the different pools and associated devices, for performance reasons or for migration purposes.

4.10.2.4 Performance
Managing the location of the physical data is very important in a virtual file system. The software must be able to decide to move data to a different location if a bottleneck comes up, or even make multiple synchronous copies of data to distribute access paths. This requires sophisticated performance monitoring and automatic tuning actions. Since the purpose of virtualization is also to give administrators an easier job, tools should be available, allowing them to visualize and monitor the overall file system, yet at the same time some artificial intelligence in the virtualization system could automatically perform the necessary actions to solve performance problems.

4.10.2.5 Security
Access to data should be protected by a security system that controls access at a file level. This requires serious consideration, since the servers will be heterogeneous and different security Access Control Lists exist for UNIX and NT, for instance. The virtualization should include global security independent of the clients' operating systems.

4.11 Summary

The nature of the data we put on computers these days has changed. Computer systems do not just run your payroll, keep your inventory or give you a set of graphs of quarterly results. They drive your results, and your payroll, because they are your shop window and your sales desks, in today's businesses. People become knowledge workers, storing data is storing the intellectual capital of your company. Computer storage also contains much of your private data, none of which you can afford to lose.

Sharing data and making more efficient usage of information with uninterrupted access is facilitated by the introduction of storage networking, whether it is SAN or NAS. Virtualization can make them more transparent to manage and exploit.

Chapter 5. Introduction to data sharing solutions

Data sharing is the generic term usually applied to common data access from different computing platforms within an enterprise. Data sharing speaks to the need to share data between various "islands of computing". That is, data must often be shared among different servers, applications, and departments.

In Part 3, "Understanding storage management software" on page 211, we discuss various IBM solutions that are available to you today. This includes data sharing between homogeneous and heterogeneous platforms. We concentrate on discussing data sharing using storage subsystems as the basis for sharing rather than telecommunications networks. Also, we discuss some software-based data sharing solutions. Discussions are limited to IBM approaches to data sharing.

Because of the many different ways data can be shared, we shall begin here by defining a set of terminology to differentiate the types of sharing. Then we will discuss the following topics:

- 5.1, "Storage sharing" on page 59
- 5.2, "Data copy sharing" on page 62
- 5.3, "True data sharing" on page 67

We use three different categories, or levels, of data access requirements for data sharing. Figure 27 documents these different types of sharing. These categories are valid for all data sharing, not just that based on storage subsystems. The three categories of data sharing are:

- *Storage sharing* splits the storage into physical partitions that are each owned by an individual attached platform. Participants in storage sharing may be aware of the presence of the other participants, but need not be.

- *Data copy sharing* involves replicating the data through some type of copy process so another platform can access it.

- *True data sharing* refers to data in a continuous storage space that can be concurrently accessed by different platforms without any replication. This implies concurrent access to files with potentially a mixture of readers and writers. Participants need to be aware of the presence of other sharers to ensure data integrity. This is usually implemented by some kind of locking mechanism.

We also use the terms *homogeneous* and *heterogeneous*. They simply mean that the server platforms doing any of the three levels of sharing are the same (homogeneous), or different (heterogeneous). For example, two S/390s running OS/390 are homogeneous platforms; while HP UX on an HP 9000, and OS/390 on a S/390, are heterogeneous.

The final goal of data sharing is to arrive at a solution that allows true data sharing between heterogeneous hosts. This goal has yet to be reached. Currently, there are solutions that allow sequential access to a single copy of data from heterogeneous platforms; however, there is currently no storage-based solution available that allows concurrent file access.
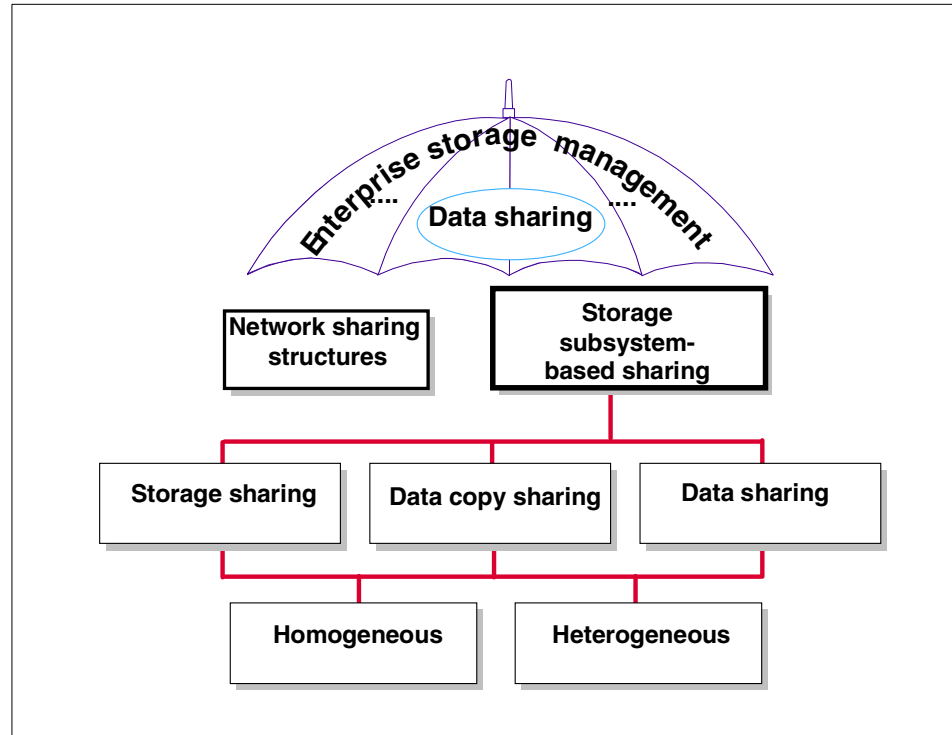
Figure 27. Relationship between the different types of data sharing

Currently, there are many solutions available in both the storage partitioning and the data copy categories. Each solution provides its own benefits. Any proposed solution needs to be evaluated in light of the environment in which it will be implemented. In the final analysis, the nature of the data and the way that it is used must be the driving force in determining the way that data will be shared.

There is a growing trend to use data sharing technologies based on storage subsystems rather than those based on the network. Some of the reasons for this are:

- The network loads are increasing faster than network capacity and network data sharing is a key network user. The network needs to be free to handle transaction processing.

- The processor overhead of using the network really interferes with processing other applications on both sides of the transfer (especially TCP/IP). TCP/IP can consume over 10% of total processor resource for network data sharing at some sites.

- Networks are almost certain to be too slow if they are affordable or too expensive if they are fast enough.

In the remainder of this chapter we describe the types of data sharing solutions that are available.

5.1 Storage sharing

Storage sharing is physical partitioning of storage resources. The only participants are the processors and the storage subsystem. The physical storage is divided into fixed partitions that are available to a given processor, so sharing is a static condition that does not require any software knowledge. Storage sharing is as applicable to tape as to disk storage. A good example of storage sharing is the IBM Enterprise Server solution. See 7.1, "The IBM Enterprise Storage Server (ESS)" on page 121, for details.

Knowledge of which processors own what storage resources is in the subsystem controller. Because the storage controller manages partitioning, reconfiguring the storage between processors may require an initial microcode load (IML) of the controller. The IML will impact all systems using the subsystem, not just those whose resources are being reassigned.

When considering specific storage sharing solutions, the ease of reassigning storage capacity between partitions is important. It is desirable for a customer administrator to be able to perform dynamic reconfigurations rather than having to schedule a long outage or requiring an on-site vendor service representative to do it.

Another factor that should be considered involves any distance restrictions imposed by the chosen solution. SCSI channel lengths may play a large role here.

Figure 28 shows an example of storage sharing in a disk and tape environment. Each of the participating systems has a subset of the storage subsystem dedicated for its use.

Servers - like *or* unlike

S/390   RS/6000   AS/400   HP   Sun   PC Server

Controller

Capacity, not data is shared
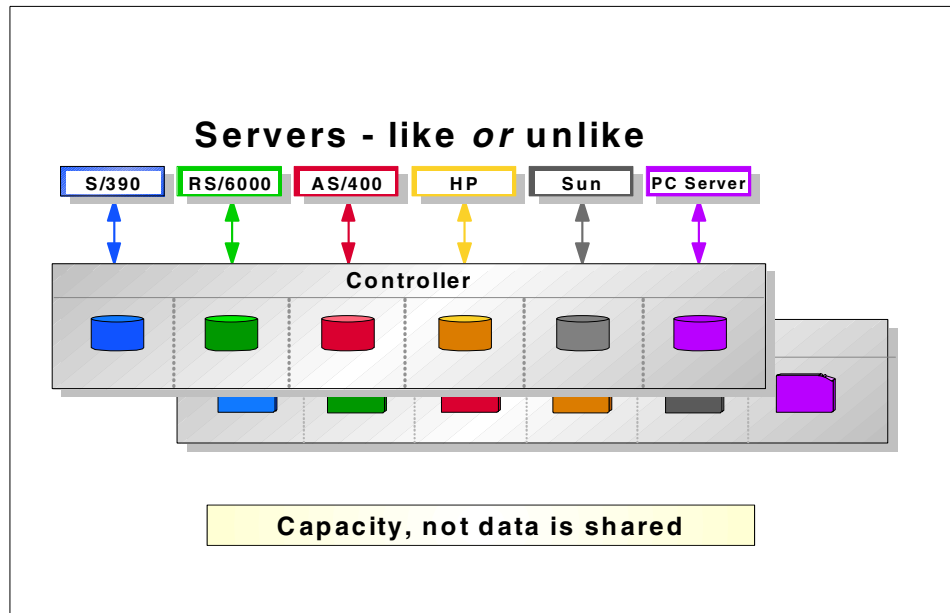
Figure 28. Storage sharing or partitioning

None of the systems illustrated in Figure 28 need to have any knowledge of the presence of the other systems sharing data. There is also no reason why some of the systems using this storage could not be taking part in other forms of storage sharing concurrently.

5.1.1 Storage partitioning goals

The major objective of storage sharing is to share physical assets with multiple servers and reduce the total cost of ownership of storage. Storage sharing gives you more flexibility in managing your storage subsystem assets, protecting and enhancing your investment in them. Because you are using a single type of storage to support more than one platform, devices can be purchased in higher quantities with potentially greater discounts. A good example of storage partitioning is the IBM Enterprise Server solution; see 7.1, "The IBM Enterprise Storage Server (ESS)" on page 121 for details.

There are also savings from having only one type of storage to administer and manage, especially because you use a single set of tools, and the storage capacity can be dynamically shifted between platforms as needs dictate. These savings apply not only to your production environment but also to any disaster recovery or contingency systems that you maintain. Storage partitioning is also appealing to companies that provide outsourced or facilities management services, because resources can easily be reassigned as customer requirements change.

5.1.2 Asset protection in storage sharing

In many ways, storage sharing provides more asset protection than data sharing. The idea is to make the disk subsystem independent of the server. As your dependence on different servers changes over time, you can use the same storage subsystems to meet the new requirements. This concept of one-box-for-multiple-platforms can also simplify disk backup. For example, if one of your servers provides superior backup and recovery functions, you may decide to use this server as a central point for backing up all other servers. If it is possible to perform storage management tasks centrally, then there are potential benefits:

- Storage management for all platforms from OS/390 makes a lot of sense because of the powerful capabilities offered in OS/390 with system-managed storage and the extended functions offered for certain storage subsystems.

- The S/390 tape library can be used across all platforms for backup. This means that a common, enterprise-wide business continuance process is followed.

- Extended functions like concurrent copy and remote copy on the 3990-6 or 9390 and SnapShot on the RAMAC Virtual Array can be used by an attached S/390 to benefit operations on other platform types that share the same storage.

Even if storage management tasks remain the responsibility of the owners of the individual systems, storage ownership needs to be centralized for shared subsystems. Even without centralizing storage management tasks, there remains the advantage of being able to reassign resources between systems dynamically.

Storage sharing requirements are usually driven by the financial executives who want asset investment protection built into the storage systems they approve. They are concerned that the I/T executive cannot forecast the demand for S/390, RS/6000, or other platform storage over the long run.

Asset protection is best provided by storage systems with the following:

- Adaptability in attachment; for example, the ability to attach both SCSI-attached and ESCON-connected devices.
- Capability to connect to heterogeneous platforms, including OS/390, UNIX, and Windows NT.

5.2 Data copy sharing

Without true data sharing solutions for heterogeneous platforms in the short term, business requirements such as data mining and data warehousing are driving the need for short-term solutions that provide faster data replication. Good data-copy sharing solutions require flexibility to deliver data when needed at the receiving system, and this should be a totally automated process. Solutions must meet application requirements by providing timely data replication between platforms of different types. There are several different approaches to data copy sharing and these are illustrated in Figure 29 on page 64, Figure 30 on page 65, and Figure 31 on page 66.

Depending on objectives, the following efficient technologies could be used:

- Snapshot (see 7.5.2.3, "SnapShot" on page 149)
- FlashCopy (see 7.1, "The IBM Enterprise Storage Server (ESS)" on page 121)
- Concurrent Copy (see A.3.2.4, "Concurrent Copy" on page 378)
- Remote copy (see A.3.2.2, "Remote Copy" on page 374)

5.2.1 Benefits of data copy sharing

The benefits from data replication fall into two major categories. The first are business benefits. These are perceived as a strategic or competitive advantage that is provided by having more current data for business intelligence applications.

The second benefits are seen within the data-processing environment. These benefits include:

- Reducing loads caused by network transfers and TCP/IP processing
- Supporting urgent business requirements
- Improving scheduling flexibility
- Improving user service
- Automating the file transfer process

No matter which of these data-copy sharing methodologies is used, if the platforms are heterogeneous, there is a very high likelihood that data is represented differently on those platforms. Therefore, data translation needs to occur in order for the receiving platform to be able to understand the data format.

5.2.2 Data copy sharing approaches

Three basic approaches have been developed for data-copy sharing based on DASD subsystems.

5.2.2.1 File transfer

File transfer was the original solution for data sharing between heterogeneous systems with solutions such as tape transfers or bulk network transfers. Today, solutions are available where transfer can occur at a storage subsystem level, and that is what we discuss here.

This method has an inherent restriction that, as a minimum, all intermediate data must reside within the same storage subsystem. Some solutions also require that all source data sets are located in the shared subsystem.

Data replication using file transfer programs is a fairly cumbersome three-step approach involving two intermediate flat files that are written to the storage subsystem as part of the transfer process. Only the second step is the actual file transfer, or use of a utility function to transfer the data to the target platform. The benefit of this approach is that existing source database extracts and the target platforms database load programs can be used with a utility. However, the process requires transfer steps involving delays and requires a lot of storage space. This approach is illustrated in Figure 29.
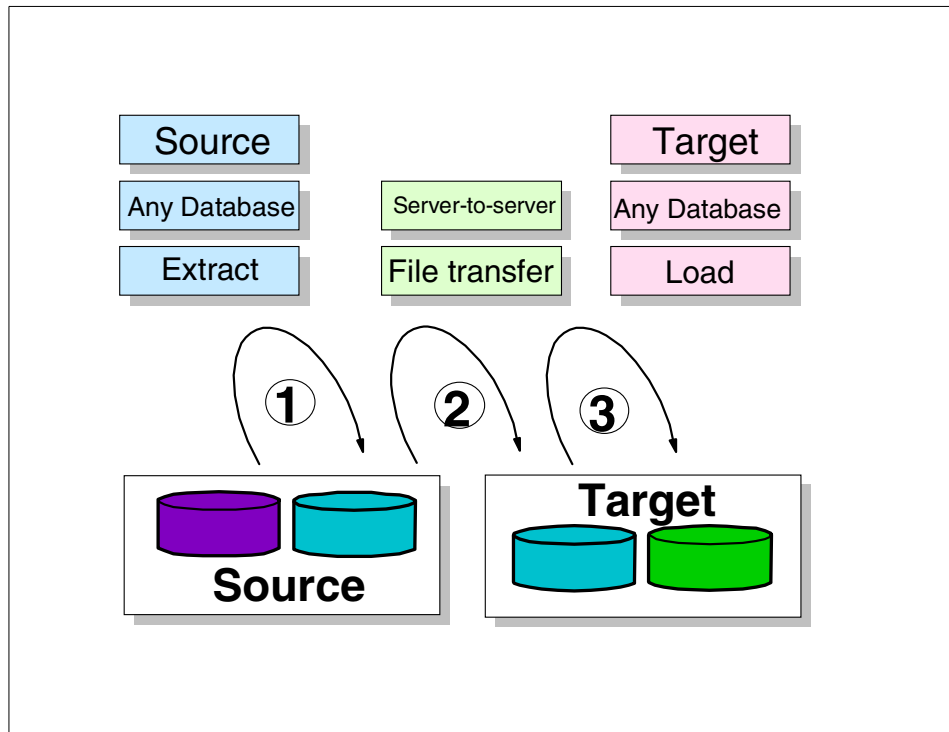
Figure 29. FTP file transfer approach

The three steps required are:

1. Extract data from the source database to a flat file on the source host storage. After extraction, the data may optionally be scrubbed to transform data to a format that is acceptable to the target system.

2. Transfer the file to the target host's storage

3. Reload the file to the database on the target host

Note that the source and target databases do not need to reside on a shared subsystem, but that the intermediate files do need to be on one subsystem if the extract process is not handled by the storage subsystem. If the storage subsystem is responsible for the extract process, then all source files, intermediate files, and potentially the target file will need to reside in the same storage subsystem.

5.2.2.2 Serial transfer approach

Using the serial transfer approach, the middle step is eliminated from the file transfer process, and the database load on the target system is from the source systems storage. This process reduces the processing required for the file transfer method, but still requires the intermediate transfer file, and still requires physical sharing of the storage subsystem.

The serial transfer approach is illustrated in Figure 30.
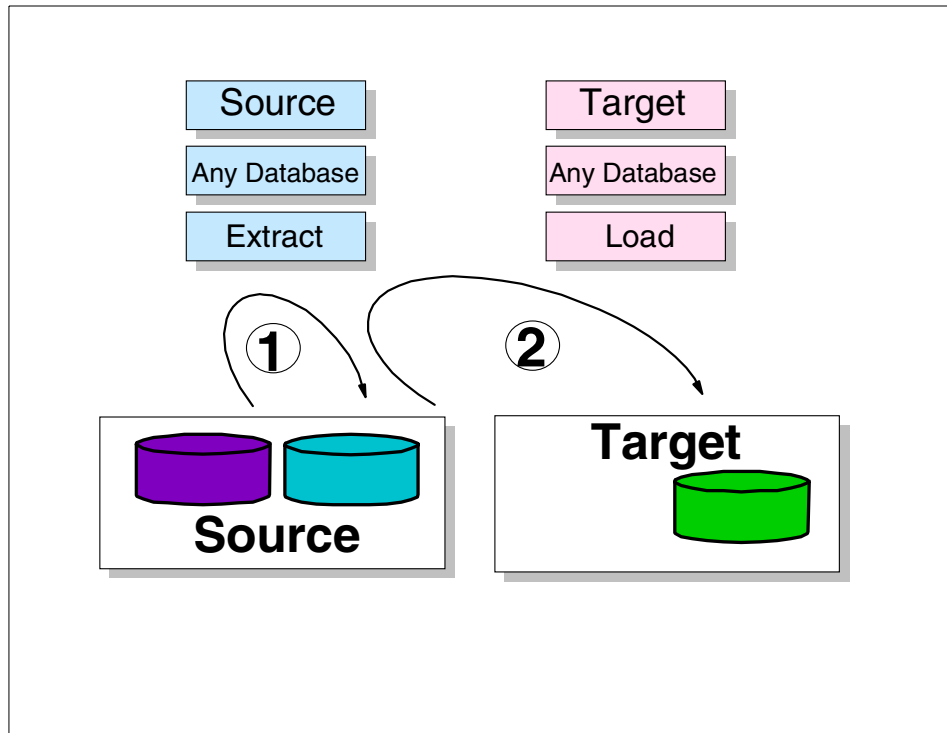
Figure 30. Serial data transfer

The two steps required for the process are:

1. Extract the data from the source database to the flat file on the source host that resides in shared storage

2. Reload the file to the database on the target host

5.2.2.3 Data piping transfer

Both of the previous data-transfer approaches require resources to be spent in creating the intermediate file. These resources included CPU cycles, I/Os generated, and space occupied. Even if this work is performed "under-the-covers" in the storage subsystems, these are still resources that are unavailable to perform productive work.

Data piping is a one-step process with no intermediate file copy. We think this approach provides the best customer value. With the fewest steps, it uses the power of the applications to collect the data from across multiple subsystems, and it uses the speed of channels (rather than networks) to transfer data.

The extract program is writing the data to a so-called *pipe* that is concurrently being read by the target systems loading process. This approach is outlined in Figure 31.
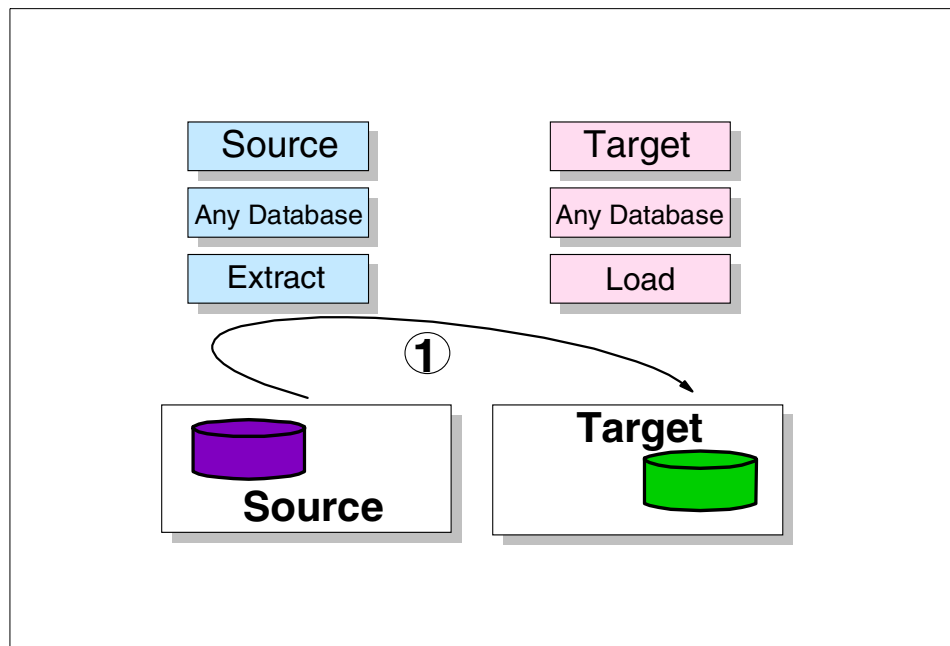
Figure 31. Data piping transfer

Since this method provides a one-step process with no intermediate flat file, it is effectively an application-to-application transfer with any required data translation, performed within the pipe. This approach does require the provision of a resource that becomes the pipe. This pipe is most often a channel with some additional hardware or software performing the data translations.

An example of the data piping approach is Client Input/Output Sockets (CLIO/S), a high-speed replication solution that IBM has available for piping data between RS/6000 AIX and OS/390 using a dedicated ESCON channel. The RS/6000 requires an ESCON adapter.

A custom API is used by CLIO/S in AIX if data is being transferred to the S/390 or in OS/390 if data is being transferred to the RS/6000. Using the API from AIX is very similar to writing data to a local tape, except that it causes the data to be written to tape or disk storage on the S/390 after it is piped to the S/390 at very high ESCON speeds. Operation in the opposite direction is similar.

Because the ESCON channel is a dedicated link between the S/390 and the UNIX/Intel platform, there is never any interference or delay when it is used, transfer is on demand, when demanded.

5.3 True data sharing

True data sharing requires simultaneous, parallel, read and write access to a single copy of data. This requires cooperation between software on the individual platforms, e-business applications, middleware including relational database managers (RDBMSs), file systems, and the operating system. All of the participating components must be involved to enable each platform to access the shared data with integrity.

Figure 32 documents some of the challenges that need to be overcome for true data sharing to be possible in a heterogeneous environment.
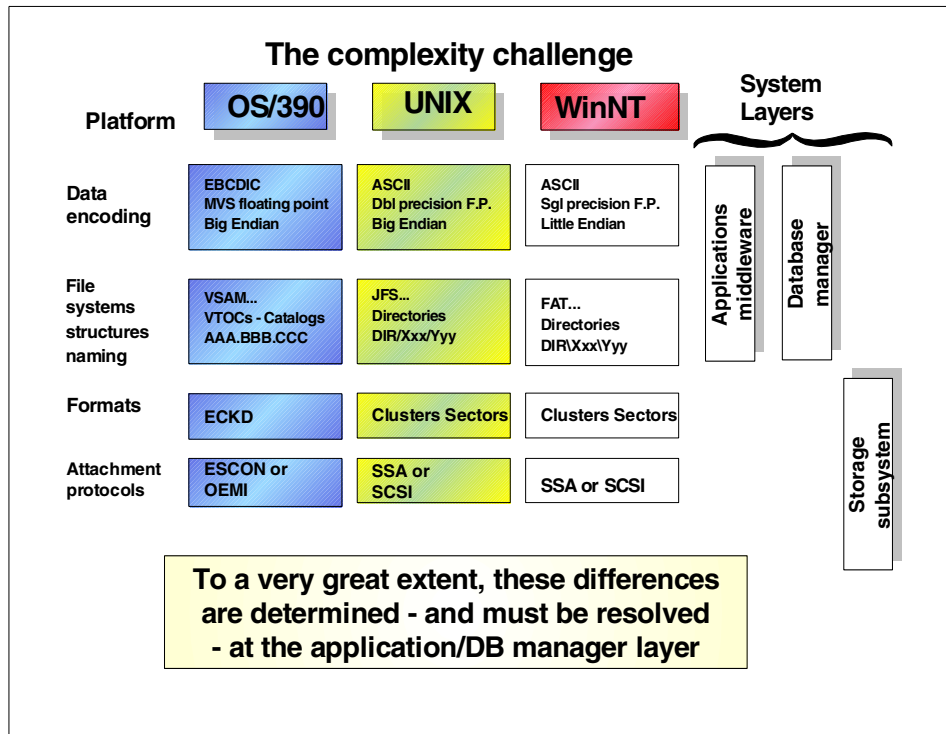
Figure 32. The challenge of data interpretation

The challenges of heterogeneous data formats are not limited to the physical format of the data, ASCII and EBCDIC; they also include the way the data is stored both logically (access methods) and physically (formats), as well as the physical attachments of devices to processors.

The intelligence supplied by the software stack is required to be able to access the data in a way that can be understood by the platform. Intelligence is also required somewhere in the systems to be able to prevent simultaneous update of the same data, since multiple system platforms all have access to the same physical copy of the data.

5.3.1 Data integrity considerations

True data sharing *must* guarantee the integrity of the data being shared, and will require code changes to the applications on at least one of the heterogeneous platforms for it to understand the data. Current DBMSs, such as DB2 and IMS, that support "true" data sharing on homogeneous platforms,

could be extended to at least partially support heterogeneous platforms, since the DBMS understands the data formats, structure, and so on.

Concurrent access to data implies either reading or writing, which leads to the danger of two different platforms trying to update the same data at the same time, an obvious integrity exposure. This means a locking mechanism of some type is required to prevent concurrent update situations. This mechanism could be implemented by software, by hardware, or by some new party introduced to play the part of an independent broker who will enable real time translation of data from one format to another.

IBM has a function called data description and conversion (DD&C) available that translates data as required for bringing it into a system that uses a different type of data representation.

Today there are instances of true data sharing that exist on homogeneous platforms — for example, data sharing in a CICSplex, or RS/6000 Scalable Power Parallel (SP2) systems running a special version of AIX (HACMP Mode 3) and a DBMS - Oracle V7 Parallel Server.

5.3.2 Benefits of true data sharing

Data sharing involves having only one physical copy of data, which can be accessed from any authorized server. This removes the need to replicate data before it can be shared, thus reducing:

- The cost of sharing data, both in duplicated media to store it and in the processing costs of generating it.

- The possibility of inconsistencies between data copies.

- The delays that occur during data replication.

- Errors introduced by either the data scrubbing or the data replication process.

A data-sharing environment is also more easily scaled; no problems appear when you add servers to the configuration. Storage management costs should also decrease, both because less data needs to be managed and because reducing or standardizing the types of subsystems that this data exists on, makes centralized storage management possible, using whichever platform provides the most appropriate tools.

5.4 Network data sharing solutions

Traditionally, data sharing has been accomplished using networks to transfer data. There are a number of different technologies that have evolved to accomplish this. These include:

- Network file servers, which keep a single copy of the data and allow read write access by clients. Examples are NFS, AFS, DFS, DFM.

- Operating system utilities, where multiple applications on the same physical platform share data. For example, OS/390 is a UNIX server.

- Relational database managers. For example, distributed relational database architecture (DRDA) provides a consistent set of rules to allow concurrent DB access, Data Joiner allows multiple DBs of different types to appear as a single database to a user, and Data Propagator allows replication of databases.

- Campus distance is a desirable feature because it will not require the relocation of servers into the glass house. (SCSI channels are limited to 25 m.)

- Distributed storage managers, such as Tivoli Storage Manager.

- Groupware such as Lotus Notes and Domino provide shared distributed data.

Networked solutions have the following advantages:

- They have a proven track record and work today.

- They are open solutions allowing data transfer between heterogeneous hosts.

- Their behavior and operation is well understood.

Network data sharing solutions that require only the transfer of required records rather than entire databases have an advantage. They move significantly less data across the network, and thus are less likely to see problems caused by limitations in network capacity.

Part 2. Understanding storage subsystems

Chapter 6. Introduction to storage subsystems

In this chapter, we explain types of storage subsystems, then discuss technologies, standards, and building blocks that are used in storage subsystems.

6.1 Types of storage subsystems

The basic types of storage subsystems available in today's marketplace are magnetic **disk**, magnetic **tape**, and **optical** disk subsystems. These subsystems have different characteristics that dictate their appropriate use.

6.1.1 Storage type selection

Selection of the disk, tape, and optical storage subsystem elements that will best suit your needs is a multidimensional process. You need to carefully examine each of the factors described below when selecting a solution. There are no simple formulas to guide you. Nearly every situation is unique, requiring careful study and analysis. There are these factors to consider:

- **Cost** — The cost of a storage system is certainly an important factor. The faster the access, the more expensive the solution. Semiconductor main storage or cache memory is very expensive, disk drives or arrays are the next most expensive, and they are followed by optical and tape drives or libraries.

    Cost per megabyte is generally inversely proportional to the capacity of the storage element. The larger disk, tape, or optical drives, like the fully populated storage subsystems, have the lowest cost per megabyte.

    In assessing and comparing costs, make sure you consider all of the cost components, including adapters, enclosures, cables, software, installation and service. It is an over-simplification to look only at the cost per megabyte of the drives or subsystem without assessing the total costs, and balancing performance against availability needs.

- **Access type** — Data transfers can be sequential or random. Sequential access means transferring large contiguous files or records. Typically, transaction logs are written as sequential files. Tape drives are best suited for sequential access and for data file backup and recovery.

    Random access implies reading or writing small amounts of data at many points on a storage device. Applications requiring random access include customer service inquiries and transaction processing programs. Disk and optical drives are better suited to random transfers, since they require less time to find randomly sized and organized records.

73

- **Data transfer rate** — The data transfer rate is the speed at which the data is written to or read from the storage device once it has been located on the device. This speed is generally expressed in MB/sec (megabytes per second) or in I/Os (input/output) per second. Both terms are often used, but they have different meanings.

  MB/sec is a measurement of the number of *bytes* being transferred per second. I/Os per second, however, refers to the number of *data blocks* being transferred per second.

  Data transfer rates depend on the storage device and on the size of records being moved. When data is streamed to a storage device as a very large file or collection of files (sequential access), maximum data rates are possible. There are fewer separate accesses to different addresses on the device. When large numbers of small random blocks of data are moved (random access), the aggregate data rates are less. More time is spent finding and accessing the data as a part of transferring the data.

  When large, sequential data transfers make up the bulk of an application, you should consider the investment in terms of megabytes per second, since relatively few blocks of data are involved.

  When many small random blocks of data are involved, however, the number of I/Os per second becomes the primary consideration.

- **Access frequency** — Access frequency refers to the number of times data on a storage device will be accessed in a given period. High access frequency implies a greater need for high reliability, minimum time to get to the data location on the storage device, and high data rates. Disk or optical devices are best suited to applications with high access-frequency requirements.

- **On-line life** — The term on-line life describes the length of time data needs to be accessible on an active storage device. Data age is usually inversely proportional to access frequency. Therefore, as the data ages and is accessed less often, it should be considered a candidate for migration to less expensive (slower) or removable media such as tape or optical.

- **Interchange or shelf storage requirements** — If the media containing the stored information needs to be exchanged with other systems or removed and stored for future use, tape or optical storage devices will be required.

- **Data longevity** — If archive, backup or reference information is to be stored for long periods of time, the integrity of the media it is stored on is important. Tape and optical media have finite shelf lives, after which

degradation or loss of the data is likely. Tapes of 4 mm, 8 mm, and 1/4-inch have shelf lives of several years, while the newer 1/2-inch tapes have a life of 10 to 20 years. Optical media has a shelf life in excess of 100 years.

- **Reliability and availability** — Reliability and availability are extremely important considerations, particularly when considering mission-critical information.

  Reliability is the ability of the storage device to function for a stated period of time.

  Availability is the ratio of the total time a storage device is capable of being used, to the total time it is required for use. Availability implies that the data will be available from a storage device when it is required. In some applications it may be acceptable to have an occasional failure that results in loss of data. Some applications can tolerate failures and limited down-time, but cannot tolerate data loss. Other applications are mission critical and can afford absolutely no data loss or interruption of service.

  Each of the storage products offers options, such as redundant adapters or controllers, to provide the proper level of data availability.

- **Regulatory requirements** — Some industries are under legal constraints governing data residing on their systems or on removable media. Restrictions can apply to access security, the length of time that records must be retained, and to what extent data must be interchangeable with other systems.

6.1.2 Making the decision

The three overlapping circles in the next three figures (Figure 33 on page 76, Figure 34 on page 77, and Figure 35 on page 79) illustrate three levels of complexity you can encounter while deciding which storage media are appropriate.

6.1.2.1 The simple case
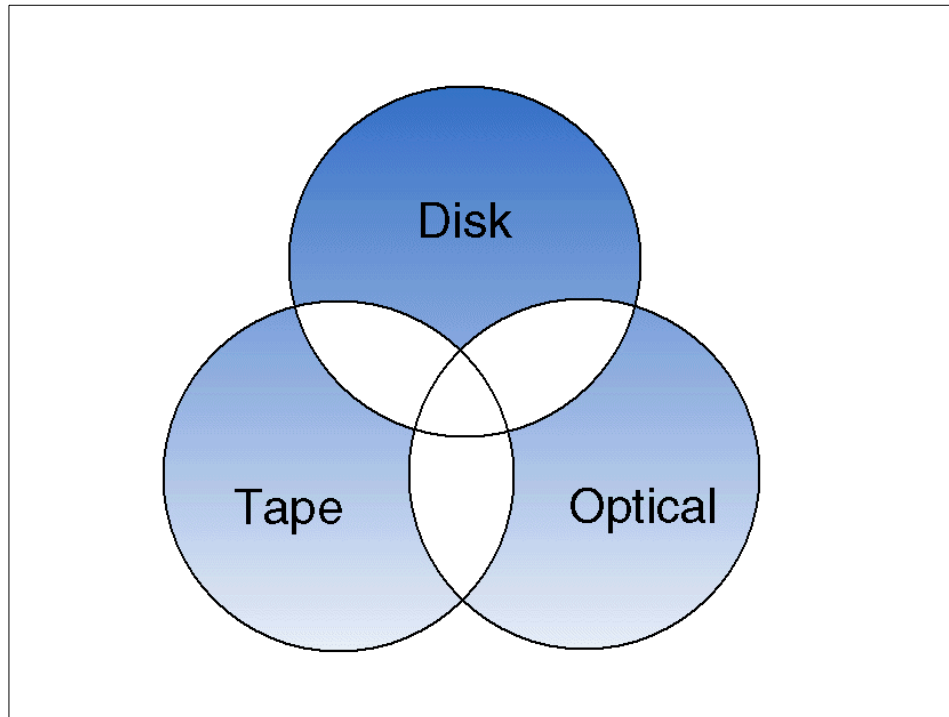
Figure 33. Simple storage media selection

Figure 33 illustrates an uncomplicated decision. Examination of 6.1.1, "Storage type selection" on page 73, clearly places the solution into one of the shaded areas:

- **Disk** — Disk is required when high performance and rapid response times are critical.

- **Tape** — Tape should be used where low cost archiving of information, high speed sequential access to data, or backup and interchange of data are required.

- **Optical** — Optical is the solution when long term archiving of information is the primary requirement.

Device selection is not always this easy. Requirements often generate a case for more than one device type.

6.1.2.2 Things become more complicated
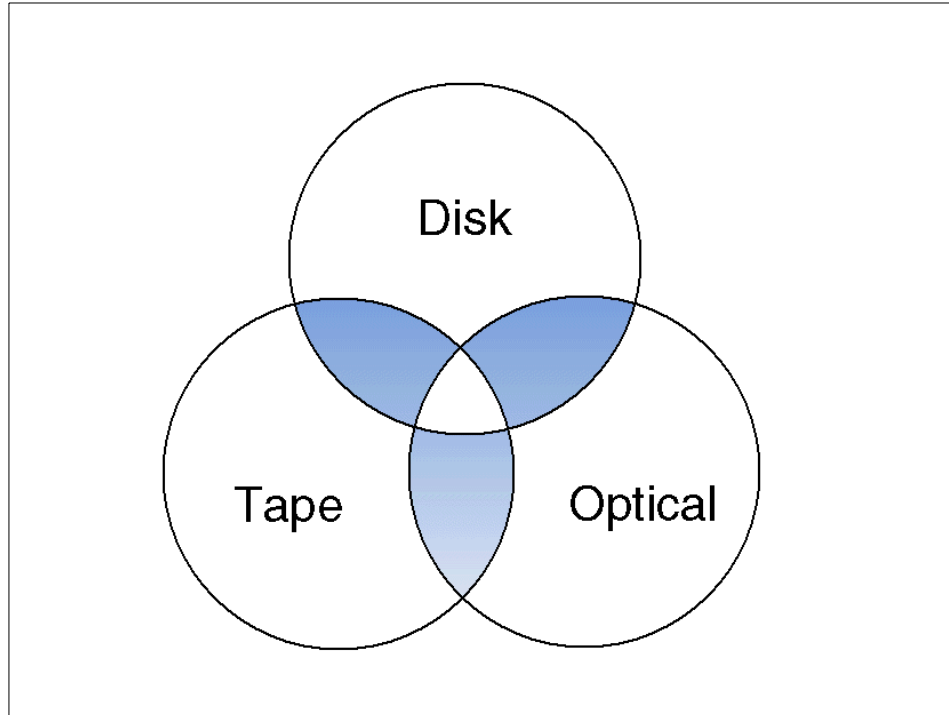
Figure 34. Requirements suggest several media types

Here the decision points resulted in requirements for more than one device type. The shaded areas in Figure 34 show where the choice between media overlaps.

- **Disk/Optical** — In this case, the requirements point to both disk and optical devices. Reasons for this include:
  - Requirements for random access with medium to high performance.
  - Nontraditional applications such as image processing indicate a need for medium access frequency to large amounts of sequential data.
  - Long-term storage of information is sometimes necessary.

- Data usage allows migration of less frequently accessed information to slower speed devices.

- High availability or fault tolerance is required.

• **Disk/Tape** — Here, the requirements include benefits from using both tape and disk. Some of the reasons for this include:

  - Access to large amounts of sequential data is needed, with medium to low performance acceptable.

  - Large amounts of space are called for.

  - The time it takes to access the first byte of data is not significant; however, the performance requirements after that first access are high.

  - Migration of less frequently accessed information from reasonably fast media to slower media is acceptable.

• **Tape/Optical** — Requirements indicating both tape and optical storage devices may have arisen as a result of the following points:

  - Most data access is low-speed sequential.

  - Random access to information is required.

  - The time it takes to access the first byte of information is not important.

  - Migration of less frequently accessed information from reasonably fast media to slower media is acceptable.

  - Data life is important.

But sometimes the decision can be even more difficult, as follows.

6.1.2.3 The more difficult case
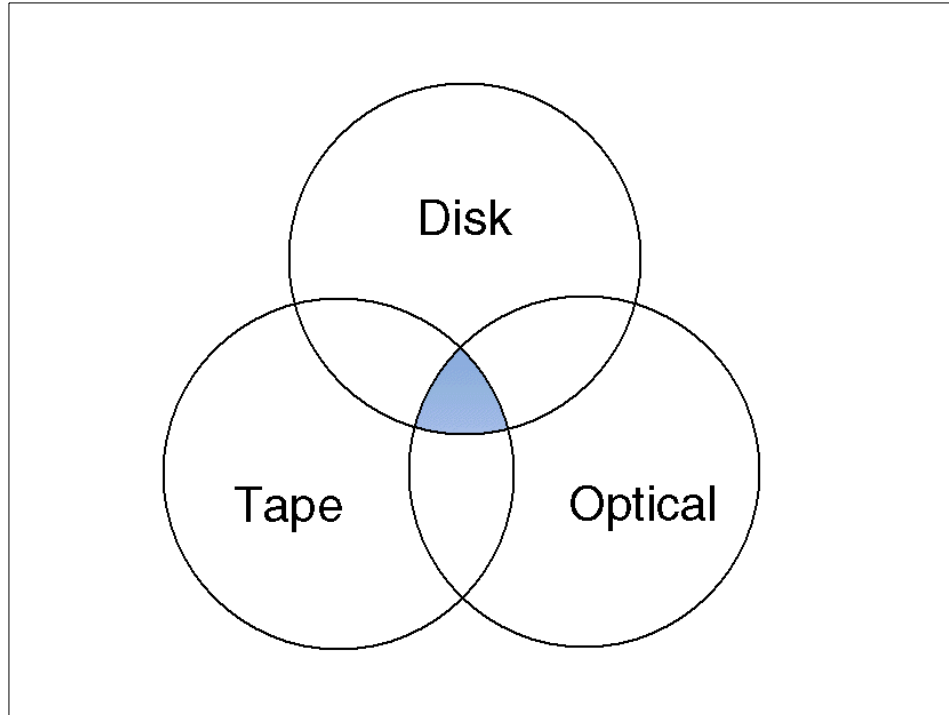
Figure 35. Complex storage component selection

The requirements are complex enough to merit selection of all three types of media devices, as in the shaded area in Figure 35. Some of the factors that may complicate the choice of media include:

- The customer demands high-performance random access to information.
- Large amounts of information need to be stored; but not all access will be concurrent.
- If storage must include archiving information, data life is an issue.
- Fault tolerance and high availability are important.
- Migration of less frequently accessed information from reasonably fast media to slower media is acceptable.
- Capacity is likely to grow.

6.1.2.4 Summary of media characteristics

Table 2 summarizes the advantages of each of the device types.

Table 2. Media characteristics

Feature	Disk	Optical	Tape
High Speed Random Access	x		
Low Cost Random Access		x	
Sequential Access Only			x
Low Cost High Capacity		x	x
Removable Media		x	x
Permanent Recording		x	

6.2 Standards, technologies, and building blocks

This section briefly explains standards, technologies, and building blocks commonly used in storage subsystems.

6.2.1 Small Computer System Interface

SCSI (pronounced "scuzzy") is the acronym for the Small Computer System Interface. Given the wide availability and generally low cost of devices that conform to the SCSI standard, SCSI has become one of the most common mechanisms in the open systems world to attach many kinds of IBM and non-IBM peripherals.

SCSI originated from the selector channel on IBM System/360 computers, and was later scaled down by the Shugart Associates Company to make a universal, intelligent disk drive interface. In 1986, SCSI became an ANSI standard and was expanded to support other types of devices as well.

This standard, now referred to as SCSI-1, allows a maximum of seven devices to be attached to the adapter and provides a one-byte parallel bus. Data can be transmitted synchronously or asynchronously, depending on the capabilities of the device being used. Both synchronous and asynchronous devices can share the same bus. Asynchronous transfer rates are typically around 1 to 2.5 MB/sec, while synchronous devices can communicate faster, from about 4 to 10 MB/sec. The original standard defined optional synchronous clock speed of up to 5 MHz, for a maximum data rate of 5 MB/sec on the one-byte bus.

While typical SCSI configurations use a single host system and a few disk and tape devices sharing a common bus, the standard does permit multiple host systems on a single bus.

6.2.1.1 SCSI-1, SCSI-2, SCSI-3

SCSI-2 is the second SCSI standard and was officially approved on January 31, 1994. IBM and many vendors have implemented the required features and most of the optional features of this standard. SCSI-2 adds a number of optional performance features, such as:

- **Command tag queuing** (**CTQ**), which provides the ability to queue multiple commands to a SCSI-2 device. This improves throughput by using more of the SCSI bus bandwidth. For example, additional SCSI commands can be passed to a disk device while it is performing a seek from a previous read or write command.

- **SCSI fast**, which permits clock speeds of 5-10 MHz but is typically implemented as 10 MHz. This implies a synchronous data transfer rate of 10 MB/sec on a 1-byte-wide (8-bit) bus and a synchronous data transfer rate of 20 MB/sec on a 2-byte-wide (16-bit) bus.

- **SCSI wide**, which defines a 2-byte-wide bus and provides a synchronous data rate of 20 MB/sec. This feature also enables up to 15 devices to be addressed on a single SCSI bus, subject to SCSI bus cabling and system software limitations.

Most SCSI-2 implementations are now 2-byte-wide and provide a maximum data rate of 20 MB/sec.

IBM's implementation of SCSI-2 is downwardly compatible with most IBM-supported SCSI devices. The original SCSI standard allowed a significant variation in the SCSI protocol implementation, and some implementations do not respond in an acceptable manner to the added SCSI-2 protocol that provides the "SCSI-fast" or "SCSI-wide" capabilities. Consequently, some SCSI devices, when attached to SCSI-2 adapters, may have problems that can result in configuration errors.

SCSI and SCSI-2 adapters and devices may be freely mixed and matched, but SCSI bus performance will be constrained by slower speed devices when the slower devices are being accessed. For example, when a SCSI-2 adapter is attached to a mixture of SCSI and SCSI-2 devices, the data transfer rate between the SCSI-2 adapter and a SCSI device will be limited to the maximum rate of the SCSI device. In this situation, the SCSI device may consume a disproportionate share of the SCSI-2 bus bandwidth, impacting the time available for data transfer to any SCSI-2 devices. In this same

example, when data is transferred between the SCSI-2 adapter and a SCSI-2 device, data transfer occurs at SCSI-2 data rates.

As another example, when a SCSI adapter is attached to a mixture of SCSI and SCSI-2 devices, data transfer between the SCSI adapter and all devices will occur at SCSI data rates, because the data transfer rate will be limited to the maximum data rate of the SCSI adapter.

As yet another example, a "wide" (16-bit) adapter can be attached to "narrow" (8-bit) devices as the result of the wide negotiations protocol defined in the SCSI-2 standard. In this case, the wide adapter will operate as a narrow adapter. Conversely, if a "narrow" adapter is attached to a "wide" device, the "wide" device will operate as a "narrow" device. Mixing "wide" and "narrow" devices on the same external SCSI bus is not recommended, due to the electrical signal variations between the first and second bytes of the SCSI bus.

The American National Standards Institute (ANSI) has currently developed SCSI-3, the next step in the evolution of SCSI. The current SCSI-3 specification centers on four interconnect technologies: SCSI-3 Parallel Interface (SPI), Fibre Channel Physical and Signaling Interface (FC-PH), IEEE P1394 High Performance Serial Bus, and Serial Storage Architecture Bus (SSA-PH).

The evolution from SCSI-2 to SPI is relatively simple: all cables, connectors, and termination essentially remain the same. SPI also uses the terms Fast-20 Narrow, Fast-20 Wide, and UltraSCSI. Fast-20 Narrow is an 8-bit bus with a maximum data transfer rate of 20 MB/sec. Fast-20 Wide is a 16-bit bus with a maximum data transfer rate of 40 MB/sec. A lot of confusion surrounds the term *UltraSCSI*. UltraSCSI is often spoken of or written in place of SCSI-3, but the two terms are not equivalent. Rather, UltraSCSI refers to the subset of SPI. In other words, UltraSCSI is another term for Fast-20 Wide; the two terms are synonyms.

The FC-PH provides high-speed serial data transmission (100 MB/sec) over significant distances. FC-PH has solved the cable-length problem of SCSI-2 (limit 6 m) by allowing data to be transferred over distances of up to 10 km. FC-PH fully supports the SCSI-3 standard, making it easier for system integrators to adopt. See 6.2.3, "Fibre Channel" on page 88 for details.

SSA-PH eases the configuration and flexibility problems of SCSI-2. Up to 128 different storage devices can be placed on the same loop of an SSA-based storage network. See 6.2.2, "Serial Storage Architecture" on page 83 for details.

6.2.1.2 Single-ended or differential SCSI bus interfaces

The SCSI and SCSI-2 standards both provide two alternative electrical SCSI bus interface configurations, referred to as *single-ended* and *differential*. These two bus interfaces are electrically incompatible, but offer different cabling flexibility. Connection of a SCSI device of one bus type to an adapter of the other type will cause the bus to be inoperable, but no permanent damage will occur because of protection mechanisms in the differential design.

- **Single-ended SCSI bus interface** — The single-ended interface consists of a single signal line for each of the SCSI data and control functions and a single ground return line for the entire set of data and control lines. This electrical interface is the simplest and supports SCSI fast data rates over a cable from 3 m to 6 m in length. The maximum length guideline for SCSI fast operation is typically 3 m unless specialized cabling and termination are used, as is the case for the IBM SCSI-2 adapter implementations. These tuned configurations extend operation of IBM-supported devices up to 6 m, provided that IBM-supported cables and terminators are used.

- **Differential SCSI bus interface** — The differential SCSI bus interface consists of two lines for each of the SCSI bus data and control functions. Implementations using this differential interface provide greater electrical noise immunity and support cable distances of up to either 19 m or 25 m at a modest increase in price. This is particularly useful to users who need flexibility in their physical layout of larger RS/6000 processors and external storage subsystems.

6.2.2 Serial Storage Architecture

Serial Storage Architecture (SSA) is a powerful high performance serial interface. It has been specifically designed for low-cost, high-performance connection to disk drives, optical drives, CD-ROMs, tape drives, printers, scanners, and other peripherals to personal computers, workstations, servers, and storage subsystems. It is the only serial interface that was designed from the outset to meet the requirements of a wide range of I/O devices.

SSA offers superior performance. Its fundamental building block is a single port capable of carrying on two 40 MB/sec conversations at once: one inbound and one outbound. An SSA connection consists of two ports capable of carrying on four simultaneous conversations, for a total bandwidth of 160 MB/sec. It uses a self-clocking code, which means that the data clock is recovered from the data signal itself rather than being transmitted as a separate signal. For transmission along copper wire, it uses the differential

pair method, requiring four wires, but it can also be transmitted along fibre optic cable.

SSA's dual-port, full-duplex architecture allows peripherals to be connected in configurations with no single point of failure. Because multiple paths are inherent in the design, increased fault tolerance is far easier to implement. SSA provides hot plugging and automatic configuration when nodes are added or deleted. For configuration flexibility, SSA nodes can be up to 25 m apart using low-cost shielded twisted pair.

SSA specifications provide the flexibility to be implemented with multiple topologies, including string, loop, and switch configurations. For example, a typical single loop could support 127 nodes (peripherals). A complex switched configuration with multiple loops could theoretically support a maximum of just over 2 million nodes.

At the same time that performance, availability, and reliability continue to improve, computing requirements are also becoming more complex. SSA's flexibility not only supports this increasing complexity, but actually permits new approaches not available with other interconnection technologies.

6.2.2.1 SSA components
Serial Storage Architecture currently comprises two components:

SSA-PH — The physical transport layer. This is the physical and electrical specification of the serial link.

SSA-SCSI —The SCSI-2 /SCSI-3 mapping. It is recognized that the logical aspects of the SCSI specification remain quite appropriate for addressing serially attached peripherals, and so this aspect of SCSI has been mapped to the physical SSA interface. This means that a transition may be achieved from SCSI to SSA with the absolute minimum of code rewrite.

6.2.2.2 SSA characteristics
The SSA-PH serial link has the following characteristics:

Topology — A flexible addressing scheme that allows connections as strings, loops, and switched loops. "Hot plugging" is also permitted.

Distance — Point-to-point connection for up to 25 m (cable). Fibre-optic connections could support distances of up to 10km between nodes with an Advanced SSA Optical Extender for flexible configuration.

Bandwidth — Full-duplex communication. 40 MB/sec in each direction. The protocol allows for speed increases, as technology permits.

Format — The unit of transmission is the frame, which can be up to 128 bytes. The minimum is 8 bytes, or about 6%.

Reliability — The link is highly reliable. The design provides considerable error detection capabilities, and a large amount of transparent error recovery.

Physical — The cables and connectors have a small form-factor.

6.2.2.3 Topology

SSA topology is a series of point-to-point links. There is no bus, and no arbitration is required. The ports are connected between the nodes with a three-way router. Up to 127 nodes can be addressed on a loop or string, all linked inbound and outbound in two directions. This point-to-point duplex linkage means that the initiating node can talk to any other device in either direction.

The SSA design allows an extremely flexible assortment of connection options. SSA networks can be connected in simple strings or loops, or more complex switch strings or cyclic paths. This flexibility allows trade-offs to be made between cost, performance, and availability. This variation is afforded by three different types of SSA nodes: single port, dual port, and switch.

Strings — A string is a simple linear network of two or more nodes. The port at either end can be single-port nodes, while the others are dual-port nodes. A special case of a string is the detached connection, where two single-port nodes connect to each other across one link.

Loops — The commonest form of connection is the loop. A loop is a cyclic network containing only dual-port nodes. Loops have the benefit of higher bandwidth (there are two data paths available between any two nodes) and higher availability than strings (any single node can fail without prohibiting communication between nodes). In addition, a node may be inserted into the loop without breaking communication.

Switches — A switch can have up to 126 ports. Switches allow large numbers of nodes to be connected together, and also enable alternative paths to be established to provide fault tolerance.

6.2.2.4 Compact cables and connectors

Each external SSA connection uses a compact shielded four-wire cable with an overall diameter of less than 6 mm. The external connectors are

microminiature D-shell with nine pins. When connections are located within an enclosure, lower cost connectors and cables, such as twisted pair and flex, can be used.

6.2.2.5 High availability and fault tolerance
Lower power requirements and fewer components mean higher reliability. Point-to-point linkage ensures excellent signal quality and correspondingly low error rates. In addition, each SSA node performs extensive error checking and provides transparent frame recovery if an error occurs.

When SSA devices are configured in a loop, alternate paths to each device result in no single point of failure. If a single link is broken, work continues unhindered because every device can still communicate with every other device. Meanwhile, the affected link alerts the initiator to the problem and isolates the failure to one unit, so repairs can be done immediately.

Since the system is self-configuring, repairs, as well as insertion and removal of devices, can be accomplished with no break in operations or loss of data.

The enhanced level of SSA availability means that RAID implementations that were not practical with SCSI are now possible. You can put data drives plus a parity drive on one loop without a single point of failure. With SCSI, of course, you would need multiple buses. The RAID levels supported are RAID 0,1,5, & 10 and can vary by platform. SSA also has a new 3-Way Copy function that be used in conjuction with the RAID 1 & 10.

6.2.2.6 Open to all
SSA technology is an open serial interface standard. Its specifications have been approved by the SSA Users Industry Group and is an ANSI X3T10.1 standard.

In addition, the SSA Industry Association (SSA IA)1 is dedicated to promoting the technical and economic benefits of SSA to the computer industry and its customers. Composed of companies from all segments of the computer industry, the SSA Industry Association is active in accelerating the adoption of SSA products and services.

6.2.2.7 High performance when it is needed
From PCs up through mainframes, SSA is the cost-effective high performance serial connection. Although other technologies may achieve greater bandwidths, the vast majority of applications do not effectively take advantage of their speed.

Data travels in small packets, each packet containing a payload of up to 128 bytes, and a highly condensed group of framing characters to handle handshaking, addressing, error detection, and spindle synchronization for disk drives. The short frame length and very low protocol overheads, coupled with the high bandwidth, ensure excellent responsiveness and throughput under high loads.

SSA's spatial reuse capability allows for higher throughput than indicated by the interface bandwidth. Spatial reuse enables separate SSA links to perform different operations simultaneously. Because each serial link is independent, you can achieve bandwidth multiplication depending on how the links are positioned relative to each other. For example, you can back up one disk drive to a tape while copying data from a CD-ROM to another disk drive and, at the same time, print a spool file to a printer. The data is transferred only between the initiator and the destination, and not sent around the loop. In effect, several initiators can be active at the same time on a common loop which multiplies the bandwidth as the configuration grows in size.

Another example of spatial reuse is shown in Figure 36. In this example, Disk 1 is transferring data at 40 MB/sec to Tape 2. At the same time, the CD-ROM is able to transfer data to Host B at 40 MB/sec, while Tape 1 is transferring data to Disk 2 at 40 MB/sec, and Disk 3 is transferring data to Host A at 40 MB/sec.
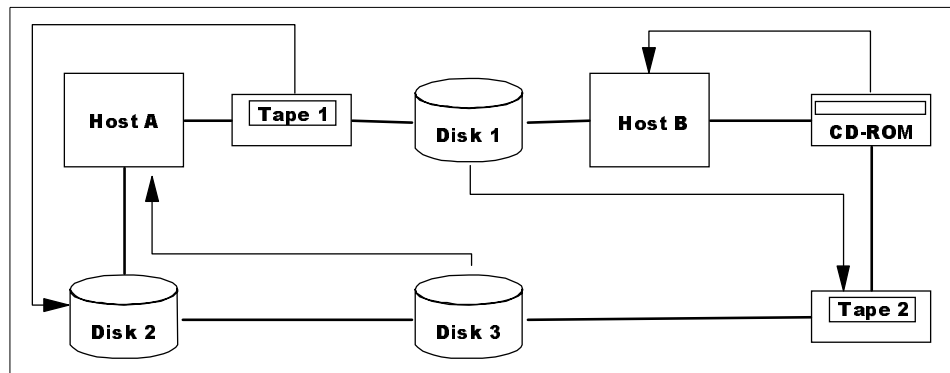
Figure 36. Example of SSA spatial reuse

> **Note**
>
> Although SSA's architecture allows transfer between devices without involving an adapter card or host, this feature is not available in today's products.

6.2.2.8 The low-cost, low-power solution

Integrated in industry-standard CMOS, SSA is the low-cost and low-power serial solution. SSA is significantly less expensive than other high performance serial alternatives and should maintain that competitive advantage as technology progresses. As for power consumption, today's SSA interface circuitry requires notably less power than other high performance serial alternatives.

SSA's design philosophy is to integrate its functions on a single silicon chip using industry standard CMOS where costs are lowest. Today, there is 80 MB/sec at the interface in 0.8 μm CMOS circuits, the current industry workhorse. When the industry moves to 0.5 μm CMOS circuits, SSA will be able to migrate there at double today's speed.

6.2.2.9 Easy migration from SCSI

SSA permits easy migration from SCSI because it is designed with two protocol levels. The lower level defines the electrical characteristics, the coding scheme, and how the information is transmitted. The upper level is a mapping of SCSI-2 commands into a form suitable for serial transmission. Consequently, device or subsystem implementations currently using SCSI-2 can be upgraded to SSA with minimal rewriting of firmware.

6.2.2.10 Flexible and Extendable

Investing in an interconnection architecture that will not meet future needs is expensive. In addition, migrating to an interim parallel approach such as a faster SCSI does not provide all the benefits of a serial architecture. SSA is the one-step approach to using the power of a serial interface today.

6.2.3 Fibre Channel

The ANSI Fibre Channel Standard had its beginnings about 1988-89. The initial motivation for its development came from I/O designers interested in higher bandwidth I/O channels that operated efficiently at fiber optic distances — tens of kilometers. The early 1990s also saw the rapid development of clustering concepts, a shift to the client-server model of

computing, emergence of high speed communications facilities and high performance devices.

The Fibre Channel Standard, originally developed by IBM, defines a high-speed data transfer interface that can be used to connect workstations, mainframes, super computers, storage devices and displays. The standard addresses the need for fast transfers of large volumes of information and could relieve system manufacturers from the burden of supporting the variety of channels and network currently in place, as it provides one standard for networking, storage, and data transfers.

According to the Standards Committee, Fibre Channel (FC) is:

- A logically bidirectional point-to-point serial data channel, structured for performance capability

- Physically, the FC can be an interconnection of multiple communication points, called *N_ports*, interconnected by a switching network, called a *fabric*, or a *point-to-point link*

- Fibre is a general term used to cover all physical media types supported by the FC and can include:

    - Optical fibre
    - Coaxial cable
    - Twisted pair

- An interface that can support multiple physical interface types and speeds

- A high speed interface that can transfer large amounts and various type of information: DATA, video, voice

- An interface that will support many different transfer mechanisms:

    - ATM/AAL (Asynchronous Transfer Method/ATM Adaptation Layer), an optimized transfer method that can handle voice communication through isochronous transfers
    - IEEE 802.2, the Logical Link Control with FC references
    - SONET (Synchronous Optical Network)
    - PI
    - HIPPIE
    - IPI-3
    - ESCON
    - FICON
    - SCSI
    - And almost anything else

- An active and intelligent interface with login and logout protocol

- A full-duplex interface which has separate fibres for each direction and separate state machines for transmitting and receiving information
- An interface that uses the 8b/10b encode/decode scheme from IBM
- An interface with different classes of transfers:
    - Class 1 Dedicated connection
    - Class 2 Frame multiplexed
    - Class 3 Datagram (ship and pray)
    - Class 4 Lossy voice and video
- An interface with different optical transmission options:
    - Single mode: allows for link budgets of 2 km and 10 km using a Class 1 laser
    - Multimode: allows for link budgets from 2 m, 1 km, 2 km, and 5 km using a 1300 nanometer (nm) LED or a 780 nm CD laser.

6.2.3.1 Fibre Channel levels

The Fibre Channel consists of numbered levels (a total of four) that interact and provide services to each other. The first few levels consist of the physical interconnection and electrical signals, as well as the rules for transmission (FC 0, 1, 2, and 3). Above the physical level exists a level consisting of an interface protocol such as SCSI-3, IPI, HIPPI, ATM, or the FC protocol controls. All data interchange originates and completes at this level.

FC-4 is the highest level in the Fibre Channel standards set. It defines the mapping between the lower levels of fibre channel and the upper level protocols (ULPs).

FC-3 defines a set of services that are common across multiple ports of a node.

FC-2 defines the signaling protocol, which includes the frame structure and byte sequences.

FC-1 defines the transmission protocol, which includes the serial encoding, decoding, and error control.

FC-0 defines the physical portions of the Fibre Channel, including the fibre, connectors, and optical and electrical parameters for a variety of data rates and physical media.

6.2.3.2 Topologies

There are three topologies that can be implemented with Fibre Channel Standard (FCS).

The point-to-point topology shown in Figure 37 is the most basic configuration in which the communication between two N_Ports occurs without the use of fabric.
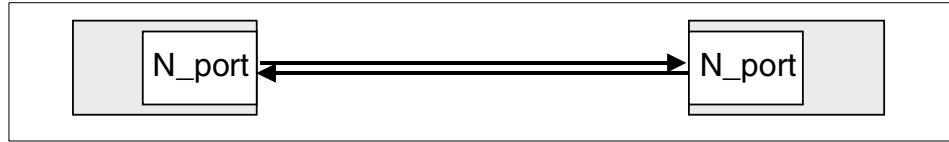
Figure 37. Point-to-point topology

The link connecting points is bidirectional and full duplex. It is a dedicated connection, meaning there are no other destinations for frames. It allows full utilization of the bandwidth.

The fabric (switched) topology shown in Figure 38 is the most complex one because of the use of fabric. This topology uses the destination identity embedded in the frame header to route the frame through fabric to the desired destination N_port. It is also the most expensive topology to implement in the FCS.
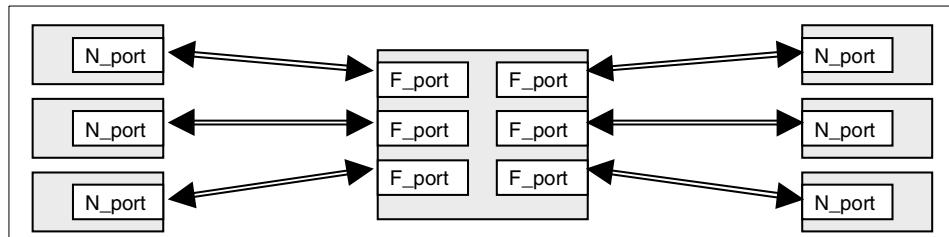
Figure 38. Fabric topology

The Fibre Channel Arbitrated-Loop (FC-AL) shown in Figure 39 is another of the topologies of FCS. The FC-AL is a shared loop technology where pairs of systems can use the loop in a time-sequenced manner. FC-AL is a serial technology with arbitration.

The arbitrated loop supports a maximum of one point-to-point circuit at a time. To establish this type of link, a single port arbitrates for access to the entire loop. Once access is obtained, it opens communication with another port. It then transmits FC traffic at full bandwidth (as if a full point-to-point

connection); other ports continue in a monitoring state at this time. Once transmission is done, it closes the connection while releasing the loop for another port. The flow of the frame is unidirectional.
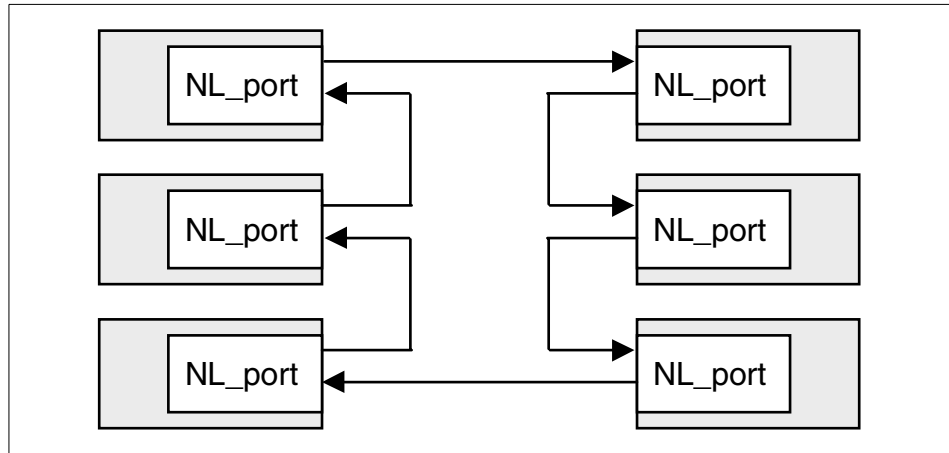
Figure 39. Fibre-Channel Arbitrated Loop topology

The FC-AL characteristics are these:

- Supports up to 127 nodes
- Has its own standard
- Can be used in combination with fabric (switched) networks
- Two operational modes
- Private loop (only NL_Ports)
- Public loop (connected to fabric switch)
- All devices share a bus bandwidth
- Requires arbitration
- Usually implemented in dual-loops
- Requires external bypass circuitry
- Provides access fairness

A port relinquishes the loop and does not arbitrate again until all other arbitrating ports have had a turn on the loop (honor system).

6.2.4 Enterprise System Connection (ESCON)

Enterprise System Connection (ESCON) is a host-device interface commonly used in S/390 platform. ESCON is the first storage area network (SAN) that IBM brought to the market.

Enterprise Systems Connection (ESCON) is the most significant change to the I/O area of data processing since the original parallel Original Equipment Manufacturer Interface (OEMI) was introduced nearly 30 years ago.

In this section, we describe ESCON terminology, fiber-optic technology, ESCON architecture and standards.

The following is a summary of ESCON announcements since September 5, 1990.

6.2.4.1 Proven ESA/390 and proven I/O interface

ESCON is an integral part of the Enterprise Systems Architecture/390 (ESA/390). ESCON replaces the previous /360 and /370 parallel OEMI with the ESCON I/O interface supporting new media, and providing proven interface protocols. Replacing the previous bus-and-tag cables and their multiple data and control lines, ESCON provides bidirectional serial bit transmission, in which the communication protocol is implemented through sequences of special characters and through formatted frames of characters.

6.2.4.2 Fiber optics

In contrast to the previous copper cables used in the parallel OEMI, ESCON utilizes fiber optic cables for data transmission. These cables are a hundred times lighter than the old bus-and-tag cables, have substantially reduced bulk, have fewer pin connections (and are therefore more reliable), have less loss and distortion, and are free from signal skew.

6.2.4.3 Topology

ESCON introduces a new topology of control unit and channel attachment. ESCON control units and channels are attached in a switched point-to-point arrangement. The switching capabilities allow multiple connections between channels and control units without requiring permanent physical connections. The point-to-point connections allow physical changes to the I/O configuration concurrently with normal operations.

6.2.4.4 ESCON Directors

To implement the switching function of ESCON, a new class of devices was introduced. These devices, called *directors*, connect channels and control units only for the duration of an I/O operation. They are nonblocking, can switch millions of connections per second, and are the centerpiece of the ESCON topology. Apart from dynamic switching, the ESCON directors can also be used for static switching of "single user" control units among different system images.

6.2.4.5 ESCON frame process

Figure 40 shows the ESCON frame process (which is quite different from the FICON frame process shown in Figure 42 on page 98).
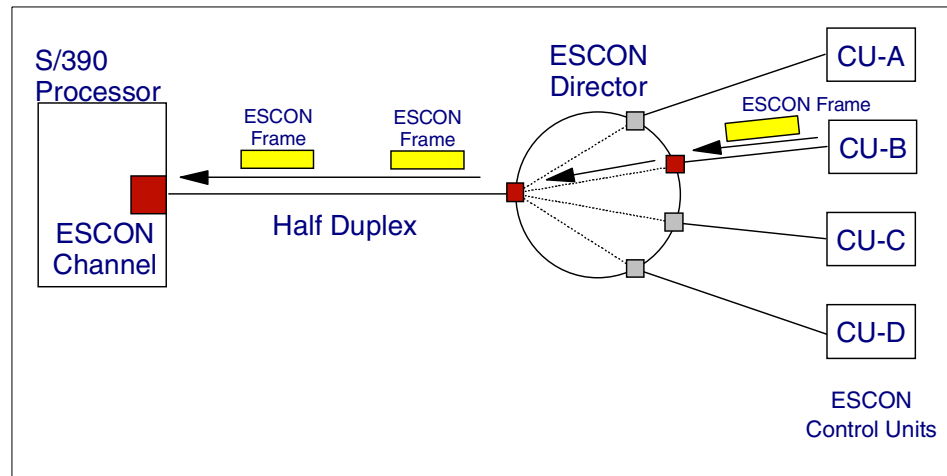
Figure 40. ESCON frame process

The main characteristics of the ESCON frame process are:

- One I/O operation at a timer per channel

- Half duplex data transfer

- Port card with 8 ESCON channels and/or control unit links, possibly to multiple systems

- Operations for other control units on the same channel wait

- Single fixed path between channel and control unit

6.2.4.6 ESCON Repeater/Converter

To allow ESCON to more easily access the common carrier trunk fiber network, the ESCON Repeater/Converter was introduced. This new device gives the common carrier the ability to monitor usage of the fiber link and also collect error statistics. These signal repeater/converters can also be used outside of the common carrier environment in place of ESCON directors where there is only a requirement for one or two fiber links to a destination.

6.2.4.7 ESCON Manager

To manage the new ESCON configuration, a new program product, the ESCON Manager, was introduced. The ESCON Manager provides an easy-to-use interface that can be automated. It offers protection from

possible disruptive changes, offers integrity through safe switching, can manage multiple systems, and is available for both MVS/ESA and VM/ESA operating system platforms.

6.2.4.8 ESCON CTC
ESCON also introduces an effective and price competitive replacement for previous channel-to-channel hardware. With ESCON channels, a user can communicate at channel speed between processors without requiring extra hardware.

6.2.4.9 Distance
Extended distances of attachment are possible with ESCON. In appropriately configured systems, control units and processors can be located up to 60 km from each other and still retain all the characteristics of control units and processors attached at much closer distances. This extended distance is exploited to provide new applications and solutions for problems such as escalating real estate prices, securing remote I/O, disaster recovery, and hot backup sites.

6.2.4.10 ESCON capable processors
A whole new processor family, the ES/9000, was announced with ESCON. All processors in this family, from the smallest to the largest, can support and exploit all the new facilities offered by ESCON. Some processor models of older families were either enhanced (ES/3090-9000 Model T) or retrofitted, such as the ES/3090 Model J, to use and support ESCON.

OS/390, MVS/ESA, VM/ESA, VSE/ESA, TPF, and the ESA versions of ACF/VTAM and MVS/DFP provide the software support for an ESCON environment.

6.2.4.11 ESCON control units
All control units available today, such as RAMAC family products and Magstar high performance tape subsystems, support ESCON attachment.

6.2.4.12 Converters
Two protocol converters were introduced to ease migration from the parallel OEMI to ESCON. They perform opposite functions, one converting parallel OEMI to ESCON protocol, the other doing the reverse. These converters are migration devices and cannot exploit the full functions of ESCON.

6.2.5 Fiber Connection (FICON)

Fiber Connection (FICON) is the latest host-device interface of the S/390 platform. It provides a greater channel and link bandwidth (100 MB/sec, as compared to 17 MB/sec for ESCON), supports up to 8 concurrent I/Os per FICON channel, intermixes control units with different channel usage characteristics on the same channel, and supports a larger number of device numbers (subchannels).

From a remote site perspective, FICON contributes to fiber consolidation, provides greater unrepeated fiber links distance (10 km or 20 km with RPQ), and moves the performance droop from 9 km to 100 km (100 km distance with repeaters).

FICON is the new I/O channel-to-control-unit architecture. It is able to ease the feeding of the constantly increasing CMOS engines of our S/390 platform servers (9672). It is also a major contributor in our cabling complexity reduction approach, making the configurations easier to manage.

This section describes the following topics:

- FICON attachment
- FICON frame process
- FICON packaging
- FICON characteristics

6.2.5.1 FICON attachment

The diagram shown in Figure 41 represents various FICON connections. The first connection, through the ESCON Director Model 5 with a FICON bridge card (see 6.2.4.4, "ESCON Directors" on page 93), permits the connection of current ESCON control units.

In the near future, IBM intends that the Enterprise Storage Server (see 7.1, "The IBM Enterprise Storage Server (ESS)" on page 121) will be able to support the native FICON direct attachment. As a result, a direct point-to-point attachment, as shown in the middle of Figure 41, would become possible, dramatically increasing the bandwidth between the control unit and the processor.

The last example, at the bottom of Figure 41, represents the intent of IBM to provide a switching capability to the FICON connectivity bringing full dynamic switching of FICON control units. This is similar to the current ESCON dynamic switching capability brought by the ESCON directors (see 6.2.4, "Enterprise System Connection (ESCON)" on page 92)
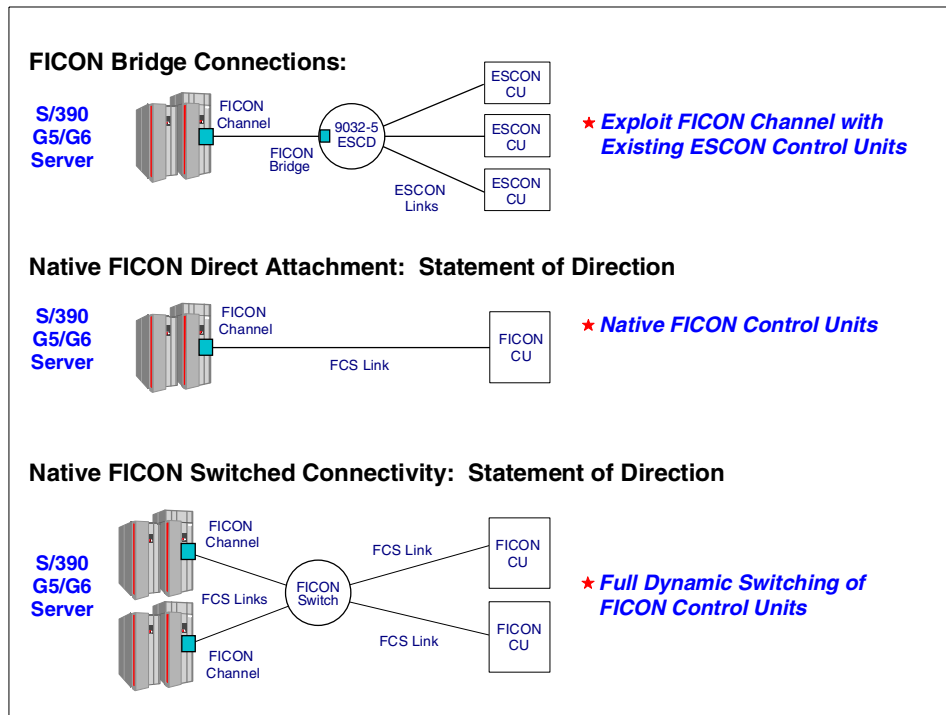
Figure 41. FICON attachment

It is also very important to understand that ESCON and FICON are not exclusive from one other, but can intermix nicely together, preserving your investment. Furthermore, FICON which basically uses single mode fibers (9 um), can reuse currently installed ESCON fibers (62.5 or 50 um cables) through Mode Conditioner Patch cables.

6.2.5.2 FICON frame process

The FICON frame process is quite different than the one from ESCON (that was shown in Figure 40 on page 94). The main characteristics of the FICON frame process are:

- Multiple I/O operations concurrently
- Multiple dynamic paths between channel and control unit
- Full duplex data transfer
- High bandwidth exploitation
- Other control unit operations run concurrently (up to 8)

Figure 42 shows features of the FICON frame process.
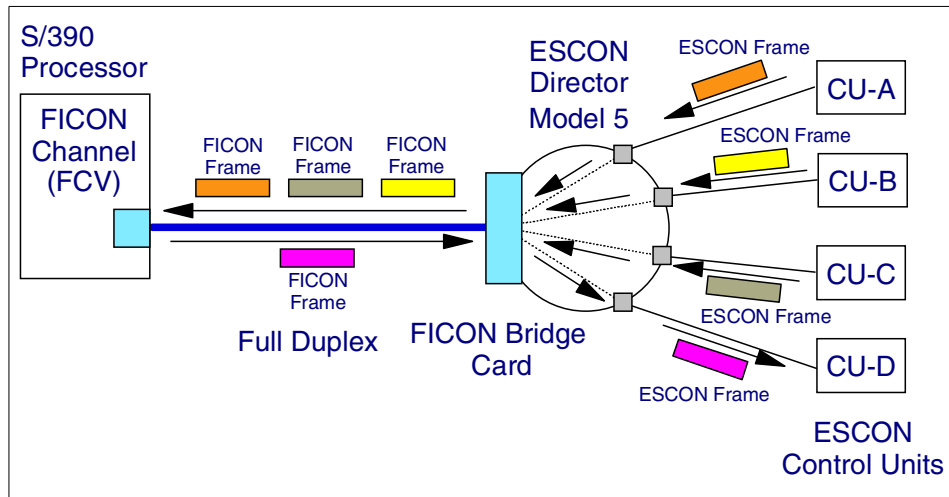
Figure 42. FICON frame process through the FICON bridge

6.2.5.3 FICON packaging

FICON attachments are located in the I/O slots of the S/390 Open Server.
Figure 43 shows the FICON channel packaging.
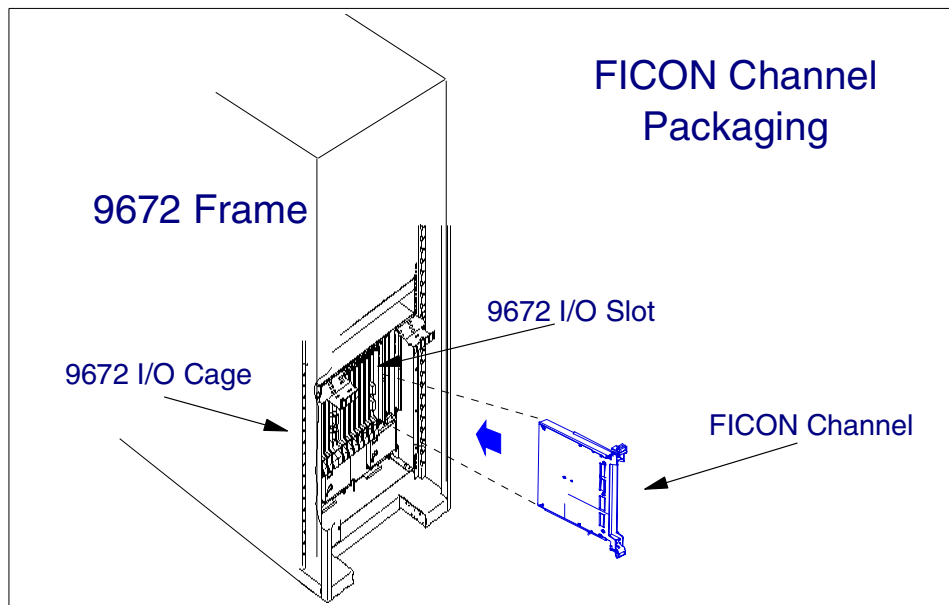
Figure 43. FICON channel packaging

6.2.5.4 FICON characteristics

The following main characteristics apply to FICON:

- Intermix with ESCON

- Bandwidth 100MB/sec

- Up to 24 FICON channels per 9672

- 1 FICON channel equivalent up to 8 ESCON channels

- Full duplex communication managing up to 8 I/O in parallel

- FICON channel supporting ESCON control units attachment through FICON bridge on an ESCON director model 5 (up to 16 bridge cards per 9032-5)

- FICON control units and FICON director will soon become available (statement of direction)

- FICON, reusing existing ESCON cabling through patch cables

- FICON, reducing cabling complexity

- Larger number of device numbers supported

- Intermixing on the same channel control units with different channel usage

- Fiber consolidation

- Performance droop moved from 9 km to 100 km

- Participates in CTC communication with ESCON CTC attachment

6.2.6 Hard disk drive

This section describes the role IBM has played in the development of hard disk drive technology in the past decade.

IBM invented the disk drive and initially led the industry with 14-inch and 10-inch disk drives for large computer systems.

In the 1980's IBM adapted its technology, first to 3.5-inch and later to 2.5-inch drives. In September 1998, we announced our industry-first one-inch microdrive technology. IBM's leading-edge hard disk drives continue to pack more information per square inch than any competitive product.

IBM manufactured the first disk drive using magnetoresistive (MR) heads in 1991, and the first disk drive using giant magnetoresistive (GMR) heads in 1997, years ahead of the rest of the industry. MR and GMR heads are able to store more information on a disk's surface because of their higher sensitivity to magnetic fields.

Areal density is the term used to measure how much information can be stored on a disk. It is calculated by multiplying the track density (number of tracks per inch) times the linear bit density (number of bits per inch on each track). Refer to Figure 44.
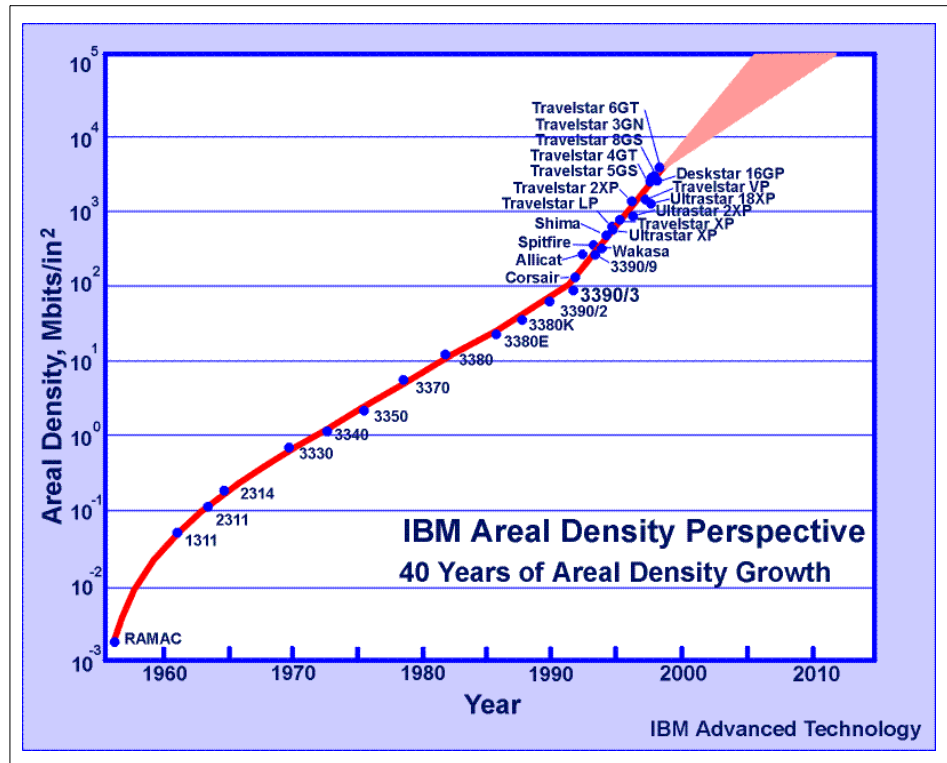
Figure 44. IBM areal density perspective

The areal density of IBM's disk drives has increased by a factor of 2 million since the first disk drive, the IBM Random Access Method of Accounting and Control (RAMAC), was introduced in 1956 (as shown in Figure 44). Since 1991, IBM disk drive areal density has grown by 60 percent per year, setting the pace for the disk drive industry.

IBM's leadership in areal density is a direct result of its leadership in MR and GMR heads.

From the late 1980s through 1999, the internal data rate of IBM hard drives increased at about 40 percent per year (Figure 45). Server class IBM hard drives have one of the highest internal data rates in the industry.
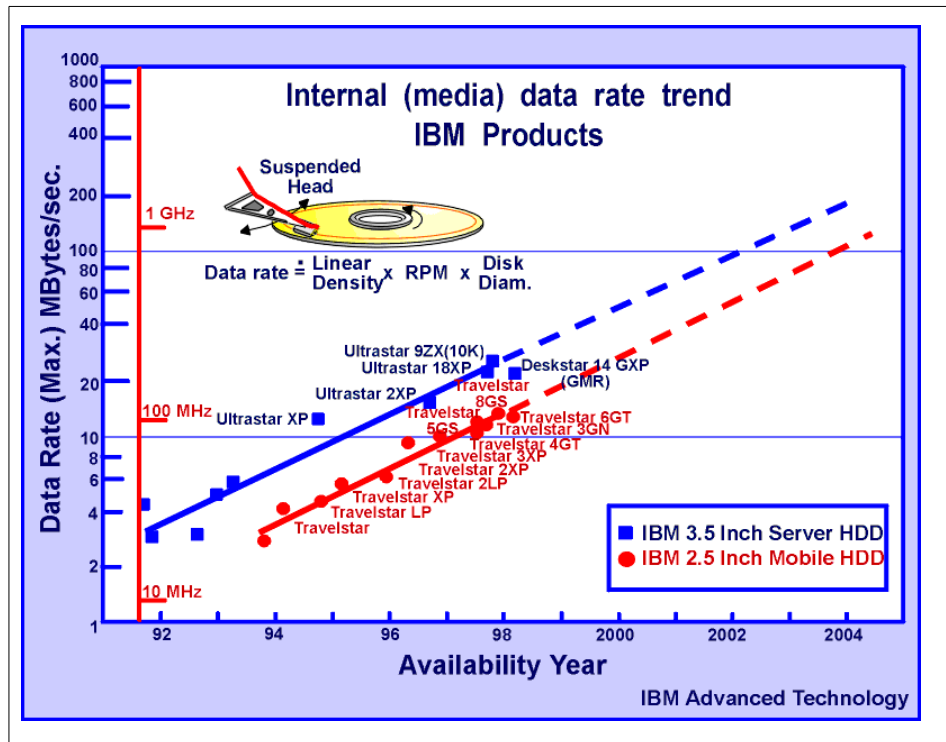
Figure 45. Internal (media) data rate trend

This increase is the result of fast disk rotation to 10,000 RPM and high areal densities. IBM has set a new world record in hard-disk data-storage density, writing and reading data bits so small that an unprecedented 20 billion of them would fit within a square inch. With a higher areal density, a drive generally is more reliable and less costly.

The GMR technology contributes to this high density. IBM has the most GMR head integration and manufacturing experience of any hard drive company. IBM was the first company to introduce a hard drive with GMR heads.

New drives IBM Ultrastar 18LZX and 36ZX are the third generation of 10,000 RPM drives offering capacities up to 36.7 GB, industry-leading average seek times as fast as 4.9 ms, the fastest seek time of any hard drive. Seek time is how fast users can find information stored on their hard drives. See Figure 46.

Figure 46. IBM Ultrastar 36ZX

These new drives are performance-optimized, with features that include a new technology, Cache Optimizer, which reduces access times to dramatically increase performance. They have buffer sizes of up to 8.0 MB, which help provide outstanding performance, especially for large data transfers and other sequential operations.

Ultrastar 18LZX can hold more than 15 symphonies in high-fidelity sound. Ultrastar 36ZX holds the equivalent of 30 movies of TV quality. Both drives are designed for use with applications that use crisp and clear audio, high-definition and life-like video, colorful graphics and animation, and enormous databases.

These drives are ideal for supporting large intranets and Internet applications, as well as digital video editing, CAD/CAM, on-line banking and credit transaction processing, digital documentation, document imaging, and data mining.

IBM's server hard drives are the only storage devices to feature Drive Temperature Indicator Processor or Drive-TIP. This allows users to monitor the temperature of the drive so it can be adjusted if the system environment gets too hot.

These drives, such as the Ultrastar unit, will be used as a building block in future Seascape-compliant subsystems.

In June 1999, IBM announced the shipping of the world's smallest hard disk drive, the IBM 340 megabyte (MB) microdrive. (See Figure 47.)

Figure 47. IBM 340 MB microdrive

With length of 1.68 inches (42.8mm) x width of 1.43 inches (36.4mm) x height of 0.19 inches (5.0mm), and a weight of 16 grams, this microdrive hard disk can hold 1000 digital photographs compressed, six hours of near CD-quality audio, 300 hefty novels, or the equivalent of more than 200 standard-size floppy disks. In one year it has since nearly tripled its capacity, up to 1GB, without increasing its dimensions

The microdrive has been designed for use in digital cameras, as well as hand-held companion and notebook PCs. With its small size and industry standard removable format, IBM's new microdrive is considered an ideal storage solution for pervasive or "go-anywhere" computing applications.

6.2.7 Redundant Arrays of Independent Disks (RAID)

RAID is an architecture designed to improve data availability by using arrays of disks in conjunction with data striping methodologies. The idea of an array — a collection of disks the system sees as a single device — has been around for a long time. In fact, IBM was doing initial development of disk arrays as early as the 1970s. In 1977, IBM filed a patent for a disk array subsystem, and was issued the patent in 1978. At that time, however, the cost of technology precluded the use of RAID in products.

In 1987, IBM co-sponsored a study by three researchers at the University of California at Berkeley on the potential use of arrays. This study, by David A. Patterson, Garth Gibson, and Randy H. Katz resulted in a paper entitled "A Case for Redundant Arrays of Inexpensive Disks (RAID)". (The name was subsequently modified by changing Inexpensive to Independent.)

The Berkeley study on arrays was a response to rapidly improving processor performance, which was increasing at 30-40% per year, while disk storage was improving at only 7% per year. The Berkeley team was looking for ways to unleash processor power by providing more data in the shortest period of time.

Patterson, Gibson, and Katz theorized that an array of small, inexpensive devices with relatively low capacity, reliability, and performance could provide improved performance by compounding the data rates of the individual devices in the array. The resulting performance would be better than that of larger, faster, more expensive, more reliable, offerings available at that time. The overall cost per unit of storage would drop because inexpensive disks were being used. Better yet, designing logical data redundancy into the array was expected to raise availability to equal or exceed that of current high-end storage products.

The original Berkeley paper emphasized performance and cost. The authors were trying to fix a performance bottleneck while lowering costs at the same time. In their efforts to improve reliability, they designed the fault tolerance and logical data redundancy which was the origin of RAID. The paper defined five RAID architectures, RAID Levels 1 through 5. Each of these architectures has its own strengths and weaknesses, and the levels do not necessarily indicate a ranking of performance, cost, or availability.

While "inexpensive" disks were the focus of the initial RAID study, subsequent work showed that large aggregates of "inexpensive" generic devices are not well suited for real applications. The word "inexpensive" in the RAID acronym was later changed by the Berkeley team to "independent," which is consistent with the fact that several independent disks in aggregate form an array.

Today, the interest in RAID is spurred by the prospects of high availability. However, there are some common *erroneous* beliefs about RAID:

- RAID makes backup of on-line data unnecessary. **Wrong**: Data must still be backed up, because when data is deleted by mistake, it cannot be recovered, even with RAID.

- RAID products never lose data. **Wrong**: While most RAID architectures can tolerate one disk failure, two disk failures could result in data loss.

- RAID products have lower performance. **Not necessarily**: Performance claims are hard to make, because performance depends on the specific RAID architecture and the type of application. RAID Level 3 can improve performance for applications accessing data in a sequential fashion, but may degrade performance when used for transaction-oriented applications. In general, RAID Level 5 has poorer performance than traditional storage subsystems.

In order to know RAID strengths and limitations, a clear understanding of the different RAID architectures is required.

6.2.7.1 RAID levels

RAID levels 1 through 5 were defined in the original Berkeley RAID paper. Subsequently, RAID levels 0 and 6 were developed.

The following four illustrations give a graphical explanation of RAID levels 0, 1, 3, and 5, the most commonly implemented RAID architectures, and the fifth illustration gives a similar explanation for RAID level 6. For convenience, RAID level 0 through RAID level 6 will be referred to as RAID 0 through RAID 6.

Parity is defined as redundant information about user data, which allows it to be regenerated in the event of a disk failure. In the following illustrations, data can mean a byte or block, not necessarily an entire file.

- **RAID 0** — Data striping, no parity

  RAID 0, also known as *data striping*, is not a true RAID architecture because there is no data redundancy. RAID 0 stripes data sequentially across multiple disks to allow parallel read or write operations, which can result in very high effective data transfer rates. All RAID 0 implementations spread the data and workload across the disks in the array, which will often give higher throughput than nonarrayed disks. However, RAID 0 does not provide redundancy along with the performance boost. As shown in Figure 48, in the event of a single disk failure, the data residing on the disk cannot be regenerated, and because of data striping, all of the striped data becomes unavailable.

RAID 0 - Striping (4 Stripes Example)

Data 1 (1/4)
Data 2 (1/4)
Data n (1/4)

Data 1 (2/4)
Data 2 (2/4)
Data n (2/4)

Data 1 (3/4)
Data 2 (3/4)
Data n (3/4)

Data 1 (4/4)
Data 2 (4/4)
Data n (4/4)

Data 1
Data 2
Data n
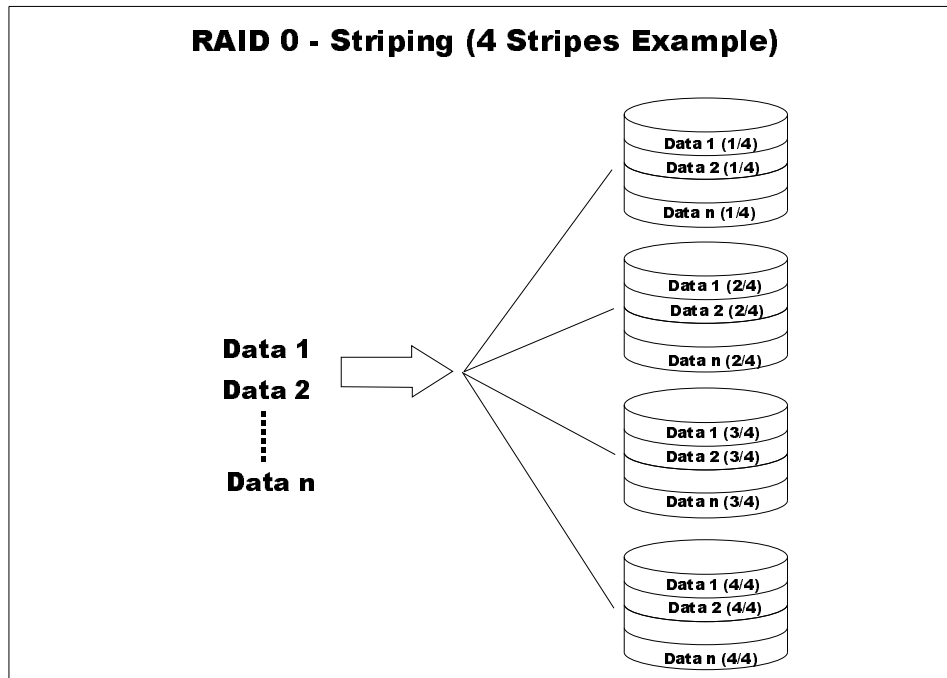
Figure 48. RAID 0

Recommendation

Never use RAID level 0 for critical applications requiring high data availability. Consider it only for applications that would benefit from the performance capabilities of this level.

- **RAID 1** — Mirroring

  The RAID 1 implementation (Figure 49) employs data mirroring to achieve redundancy. Two copies of the data are created and maintained on separate disks, each containing a mirror image of the other. RAID 1 provides an opportunity to improve performance for reads, because read requests will be directed to the mirrored copy if the primary copy is busy. RAID 1 is the most expensive of the array implementations because the data is duplicated. However, it provides the best data availability — from a disk failure standpoint — because it uses the fewest disks in its array configuration. The fewer the disks in an array, the lower the probability of multiple disk failures. In the event of a disk failure, RAID 1 provides the highest performance, because the system can switch automatically to the

mirrored disk with no impact on performance and no need to rebuild lost data.

Note

The above discussion assumes that devices with the same reliability characteristics are being compared. It also assumes that, when comparing the resulting data availability characteristics, the comparison arrays have single parity designs.
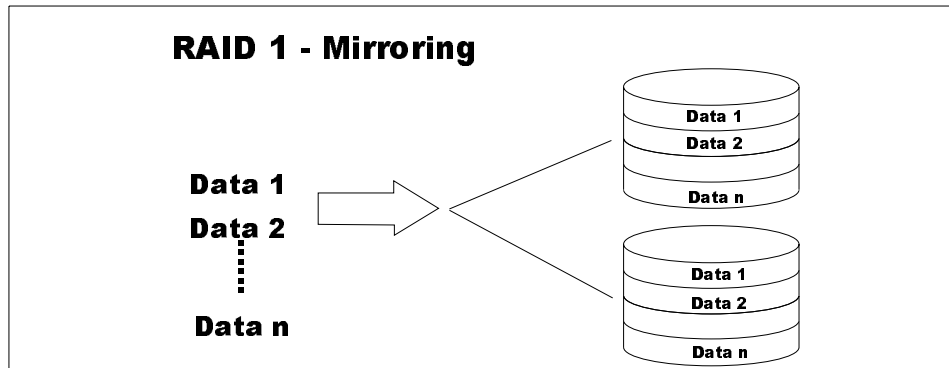
Figure 49. RAID 1

Recommendation

Consider RAID 1 when cost is not a factor for applications requiring high data availability and high performance.

- **RAID 2** — Parallel access, data striping with hamming code

  At a conceptual level, RAID 2 and RAID 3 are similar. Both RAID 2 and RAID 3 distribute data across several disks with striping at the bit or byte/multibyte level. The data is written or retrieved in one parallel movement of all of the access arms. RAID 2, however, uses an encoding technique called the Hamming error correction code to provide error detection and correction. This encoding technique requires multiple disks for the error detection and correction information, making the RAID 2 parity implementation more complex and more expensive for general use than the RAID 3 design. Therefore, RAID 2 has been of little interest in a commercial environment.

- **RAID 3** — Parallel access, data striping with parity

Unlike RAID 2, RAID 3 uses a single dedicated disk to store parity information. But like RAID 2, RAID 3 stripes or distributes data sequentially across several disks. The data is written or retrieved in one parallel movement of all of the access arms.

In the example in Figure 50, there is an array of four disks. Three of the disks are used to store data, and the fourth is used to store parity for the three data disks. If one of the data disks should fail, the parity disk can be used, along with the remaining data disks, to regenerate the data. If the parity disk fails, access to data is not affected.

Because of the single parallel movement of all access arms, only one I/O can be active in the array at any one time. Because data is striped sequentially across the disks, the parallel arm movement yields excellent transfer rates for large blocks of sequential data, but renders RAID 3 impractical for transaction processing or other high throughput applications needing random access to data. When random processing does take place, the parity disk becomes a bottleneck for write operations.
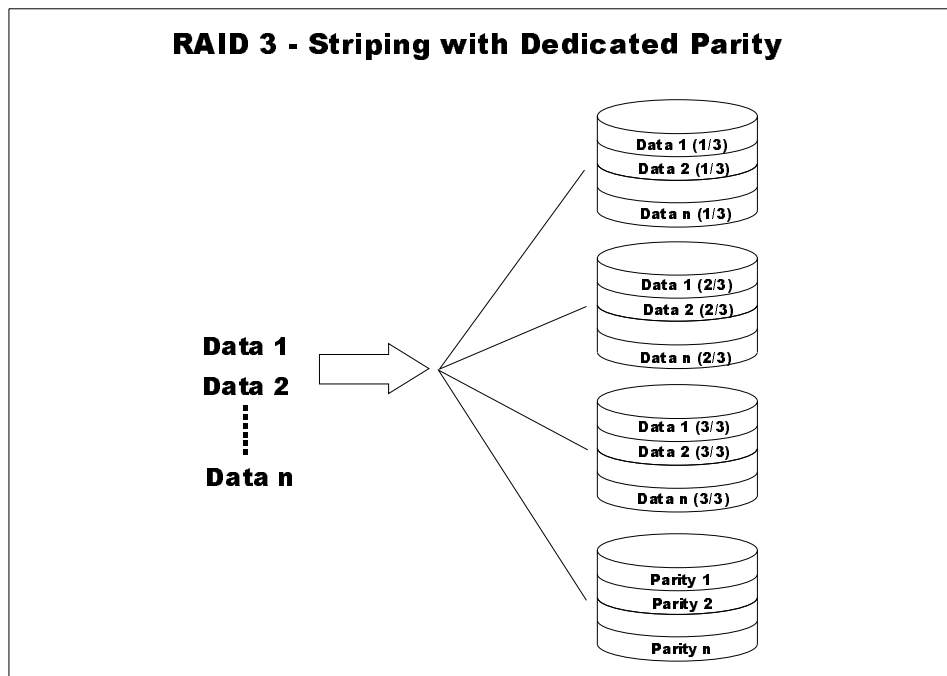
Figure 50. RAID 3

- **RAID 4** — Independent access, data striping with dedicated parity

  Like RAID 2 and RAID 3, RAID 4 and RAID 5 stripe data across several disks, but the striping increment is a block or record.

  There is only one parity disk in the RAID 4 design; in all other aspects RAID 4 is identical to RAID 5. Because the parity disk is involved in every write operation, it can become a bottleneck for transaction throughput. Therefore RAID 4 is not considered viable for commercial applications.

- **RAID 5** — Independent access, data striping with distributed parity

  RAID 5 does not have a dedicated parity disk, but instead interleaves both data and parity on all disks.

  In RAID 5 the access arms can move independently of one another (Figure 51). This enables multiple concurrent accesses to the array devices, thereby satisfying multiple concurrent I/O requests and providing higher transaction throughput. RAID 5 is best suited for random access data in small blocks.

  One important difference between RAID 3 and RAID 5 is that, in RAID 3, every transfer involves all of the disks. In RAID 5, most transfers involve only one data disk, thus allowing operations in parallel and giving higher throughput for transaction processing. In RAID 3, the segment size is one sector, and the unit of transfer is one sector for each data disk. In RAID 5, the segment size is larger, and most transfers involve only one data disk, thus allowing operations in parallel and giving higher throughput.

  There is a "write penalty" associated with RAID 5. Every write I/O will result in four actual I/O operations, two to read the old data and parity, and two to write the new data and parity.

RAID 5 - Striping with Distributed Parity

Data 1
Data 2
Data 3
Data 4
Data 5
Data 6
Data 7
Data 8
Data 9
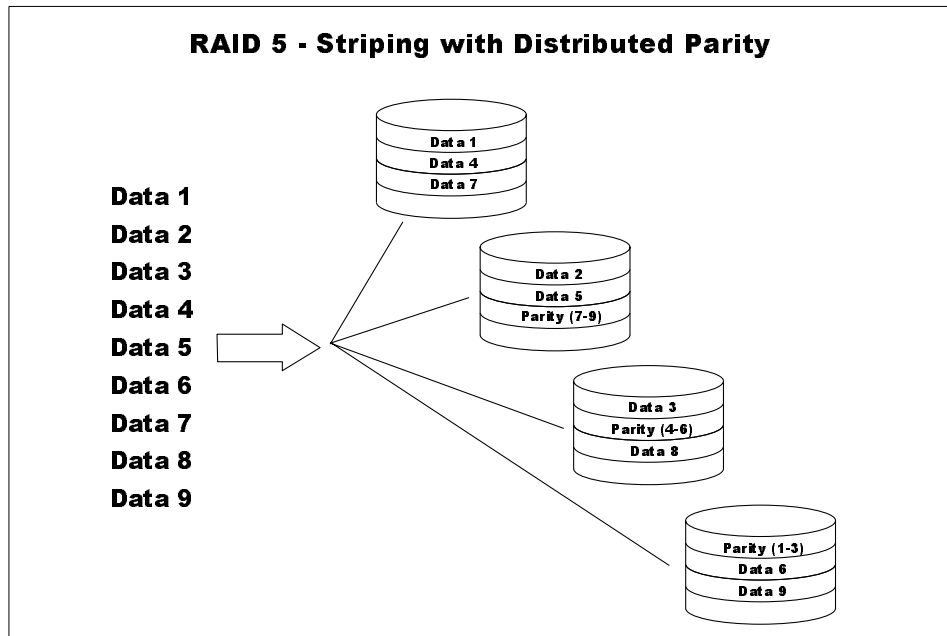
Figure 51. RAID 5

> **Recommendation**
>
> RAID 5 should be considered for environments requiring high data availability and with applications that process relatively short data records or a mixture of large sequential records and short random blocks.

- **RAID 6** — Independent access, data striping with double distributed parity

  RAID 6, which was not among the original Berkeley RAID levels, adds a second, independent parity block to RAID 5. Figure 52 shows an example of RAID 6 with an array of six disks. Four disks are used for data, and two disks are used for parity, with data and parity rotating and interleaving within the array. With two independent parity schemes, each using a different algorithm, data availability is extremely good, and is uninterrupted even when two disk failures occur at the same time. However, more disk space is required for parity, and there is an even greater write penalty than with RAID 5. For this reason, the write performance of RAID 6 is extremely low. The lower performance and the complexity of implementation have made RAID 6 impractical for most applications.

RAID 6 - Striping with Double Distributed Parity

Data 1
Data 2
Data 3
Data 4
Data 5
Data 6
Data 7
Data 8
Data 9
Data 10
Data 11
Data 12

Data 1
Data 5
Data 9

Data 2
Data 6
Data 10

Data 3
Data 7
Parity (9-12)

Data 4
Parity (5-8)
Parity (9-12)

Parity (1-4)
Parity (5-8)
Data 11

Parity (1-4)
Data 8
Data 12
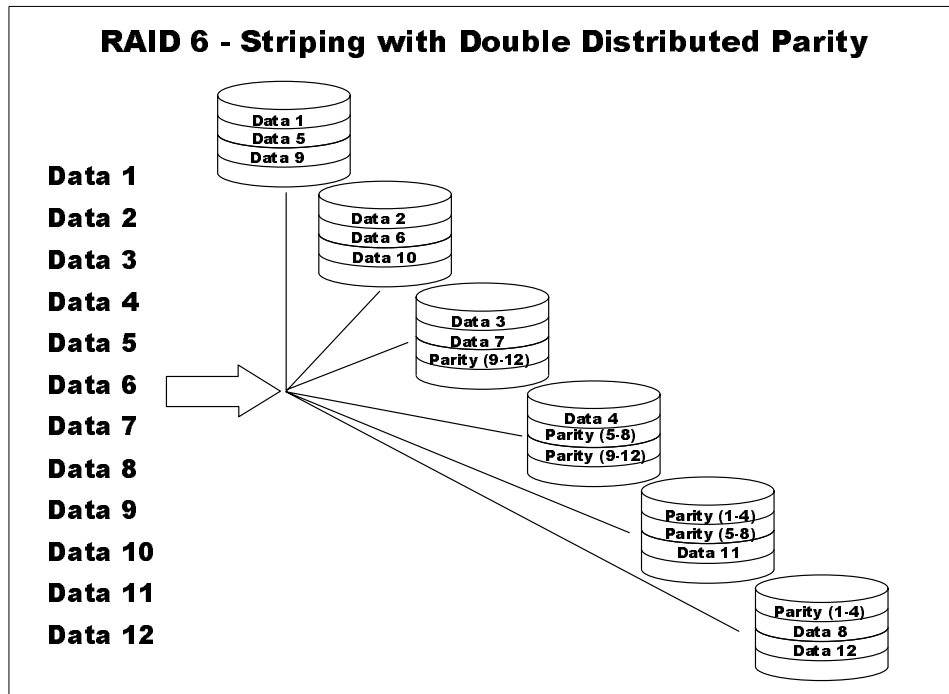
Figure 52. RAID 6

6.2.7.2 Reducing the write penalty

One way to reduce the RAID 5 write penalty is by using cache, which allows increased performance by temporarily storing data in anticipation of I/O requests to disk. By retaining accessed data in cache, read requests for the same information can be quickly satisfied from cache without additional I/O operations to disk. This can reduce disk I/O because of the locality of reference in access patterns. Cache can be used to bundle write requests until a large block of contiguous data can be written to disk in a single I/O operation. During write operations, data is first written to cache, allowing a transaction to complete sooner than the actual write operation to disk. By allowing the application to proceed in parallel with actual physical I/O operations, cache can reduce the delay associated with reading or writing data and improve RAID 5 performance.

Nonvolatile storage (NVS) keeps data stored in cache from being lost when a power outage occurs. NVS is the cache that switches to battery power during an outage. Under normal circumstances, NVS functions in much the same manner as regular cache. As soon as a write request is completed to NVS, the application can proceed without further delay. However, if an outage does occur, the updated data is still safely stored in NVS until that data can be written to the disk.

6.2.7.3 Trends in RAID

Many vendors incorporate the acronym RAID in product names or descriptions.

RAID 5+, RAID 6+, RAID 7, RAID 10, RAID 53, RAID-S, AutoRAID, and Database RAID have all been used. Typically, such products either blend multiple Berkeley RAID levels together, or add additional features to a storage system that implements one or more of the Berkeley RAID levels.

The RAID Advisory Board (RAB), an association of suppliers and consumers of RAID-related products, as well as other organizations with an interest in RAID Technology has developed Storage System Classification Program based on number of criteria. RAB believes that the information provided by classifying a storage system in accordance with the degree to which it exhibits Extended Data Availability and Protection (EDAP) attributes is more useful to end-users involved in procuring storage systems than knowing which RAID levels a product supports. RAID levels basically describe how data and redundant data are mapped across the disks of an array; information which is less important than an understanding of just how much EDAP is provided by a storage system.

6.2.7.4 Summary

In order to determine which RAID level to use in a situation you need to know the customer's storage environment and application characteristics. Table 3 helps you understand and compare the attributes of each RAID level.

Table 3. Comparison of RAID levels 0, 1, 3, and 5

	RAID 0	**RAID 1**	**RAID 3**	**RAID 5**
Method Used	Disk striping	Disk level mirroring	Parallel transfer disks with parity	Independent data access
Disks Required	n	2n	n+1	n+1
Data Protection	None	Very High	High	High
Data Rate	Very high	For Read: 2 X single disk For Write, similar to single disk	Similar to single disk	Similar to single disk
Performance	High	High for read-intensive Medium for write-intensive	Medium	Medium (possible write penalty in write-intensive environment)
Cost	Low	High (up to twice RAID 0 cost)	Medium	Medium

Notes

- RAID 0 — Each data block or record is subdivided across the disks in the array. No data protection is provided (parity or redundancy).
- RAID 1 — All data is duplicated. A copy of the data is kept on separate disk drives.
- RAID 3 — Each data block or record is subdivided and distributed across the data disks. Parity for the data is mathematically created and stored on a dedicated parity disk.
- RAID 5 — Data blocks or records are distributed as blocks or records, and not subdivided as in RAID 3. Parity is interspersed with user data.

6.2.8 Magstar and Magstar MP Tape Drive

Magstar and Magstar Tape Drive are used as building blocks of several Seascape-compliant storage servers such as Network Storage Manager, Web Cache Manager, and Virtual Storage Server. See 8.5, "3570 Magstar MP Tape Subsystem" on page 172, and 8.1, "3590 High Performance Tape Subsystem" on page 153, for technologies and specifications used in each drive.

6.2.9 Linear Tape-Open technology

At the time of writing, Hewlett-Packard Company, IBM Corporation, and Seagate Technology Inc. have just unveiled Linear Tape-Open (LTO) technology, a powerful open tape architecture, setting the stage for a new generation of tape storage products that are expected to surpass current tape capacity and performance benchmarks while maintaining the highest data integrity.

6.2.9.1 Overview

To best serve customers across the entire spectrum of applications, the three core technology providers designed LTO technology to accommodate a range of storage requirements from single server to complex network environments, in both fast-access and high-capacity formats. For this reason, two implementations of the technology were developed: Ultrium and Accelis. Each is designed to meet customer needs well into the next decade. See 6.2.9.3, "Specifications" on page 116 for further information.

LTO was also developed to take advantage of the best available servo and head technologies, which are essential to provide the high degree of accuracy critical for high track densities. The more tracks there are on a tape, the greater the risk of overwriting a neighboring track. LTO's track by following servo monitors, and control the position of the head on the tape to ensure accuracy and avoid overwrites.

The LTO Accelis format takes the Magstar MP 3570 concept to dramatically higher capacities, and to new performance levels. It will provide the same fast access and automation reliability attributes as the MP 3570, and will use the same cartridge dimensions and characteristics (self-enclosed tape path and midpoint load).

The Magstar 3590 product family will receive an enhancement plan which includes the use of the LTO technology extensions to bring the 3590 to new capacity and performance levels. These Magstar 3590 enhancements will be available to customers in the same time frame as the new LTO format.

6.2.9.2 Key features

LTO will offer a higher-recording real density (100 Mbits/sq. inch) than any linear tape technology available in the industry today. The high linear density is supported by a robust logical format, including lock step recording across multiple data tracks of user data protected by a true cross-product Error Correction Code (ECC) designed for robust multiple track operation, even if random error rates become high.

Other aspects of the robust logical format include:

- An improved data compression algorithm
- Dynamic rewrite of data written onto defective areas of a tape or written by a bad write head
- Dynamic discontinuation of writing on any region of tape with an unreliable servo region with subsequent rewrite down tape
- Use of LTO Cartridge Memory (LTO-CM) to allow old data to be easily disregarded and new write operations to be precisely triggered

LTO-CM is an electronics and interface module, embedded in both the Ultrium and Accelis cartridges (Figure 53), which contains 4 KB nonvolatile memory for storage and retrieval of information about the cartridge and data. LTO-CM also permits blocks to be recorded with an absolute Write Pass number and sequence number. This permits simple determination and disregard of old data, as well as reliable read and write operations in even the most challenging environments. It can also contain user specified information. For example, a customer may need to store manufacturer's data for warranty purposes, or data concerning the age of the cartridge, how many loads it has endured, how many temporary errors have accumulated, how much tape has been demarked, and so on.
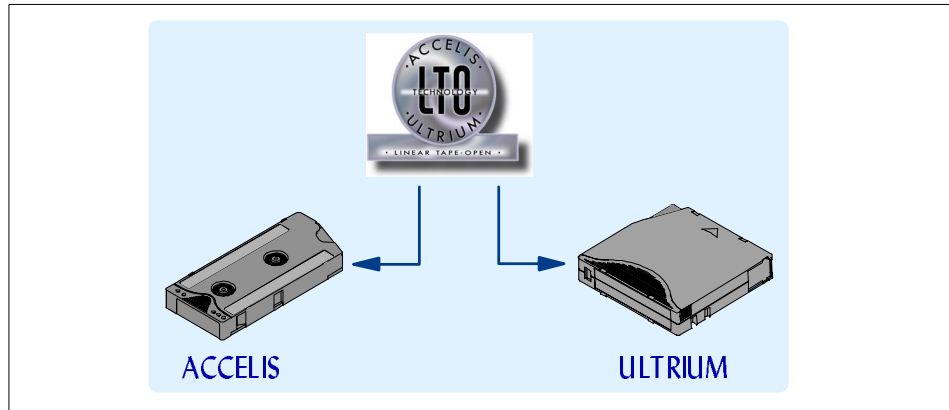
Figure 53. LTO formats

Both the Ultrium and Accelis formats share the "best-of-breed" technology provided by LTO's advanced multi-track recording capability, head and servo technologies, and high recording density. Although not specified, the characteristics of LTO were also defined to maximize the benefits of magneto-resistive (MR) head technology. The use of MR heads enables the high number of concurrent channels and the high track density characteristics of LTO.

6.2.9.3 Specifications

Apart from LTO, however, Ultrium and Accelis (shown in Figure 53) each have their own unique set of characteristics that make them well suited for their marketplace needs.

The **LTO Ultrium format** enables high-capacity, high-performance tape back-up solutions for the data storage industry. The first generation of this technology allows for storage of up to 100 GB of data (>200 GB compressed) on a single cartridge at data rates as high as 20 MB/sec (>40 MB/sec compressed).

The performance is made possible through the use of linear recording techniques that employ either a four- or eight-channel head and closed loop servo technologies. The high capacity results from applying LTO to a new data cartridge that has been designed to maximize the amount of tape surface area while still enabling very small form factors. The first generation cartridge will contain 600 meters of half-inch tape to achieve its 100GB native capacity. The new cartridge will also have a cartridge memory chip that is part of the LTO interchange specification.

The drive mechanism will provide bi-directional tape motion during read/write and locate/rewind operations. It will also include the read/write head, a mechanism to control the positioning of the head (the head actuator), sensors to detect the presence and write-protect status of the cartridge and the LTO-CM electronics and interface.

Ultrium's four- or eight-element array of read and write elements, and the extra servo read elements in the head may be interleaved (alternated in the layout), merged (stacked), or arranged in modules. The elements are arranged on the head in pairs such that a data pattern written by a write head is verified immediately by a paired read head. The heads must be capable of reading and writing while the tape is moving in the forward or the reverse direction.

Ultrium partitions the half-inch tape into 384 data tracks, evenly divided into four data bands. The number of tracks in a data band is the same whether the data is written by a four-element or eight-element head. If an eight-channel head is used, the 96-track data band results from six head positions dictated by the servo. The four-channel implementation requires extra servo readers on the head to generate 12 head positions from the six predefined positions.

Once a data band is filled, the head is aligned on the next prescribed pair of servo bands and begins to write data in the same manner. Special offsets in the servo bands ensure that the head is on the correct data band for reading or writing.

The Ultrium single-reel cartridge design uses a take-up reel that is located inside the drive. The tape is engaged by means of a coupler that "grabs" a leader pin at the start of the tape and guides it around the tape head to the take-up reel in the drive. After the leader pin is secured in the take-up reel, the reel rotates and pulls the tape through the tape path. The drive's reel motor is coupled to the cartridge by means of a clutch with a toothed rim. Figure 54 shows the expected migration path for Ultrium.

	Generation 1	Generation 2	Generation 3	Generation 4
Capacity - Native	100GB	200GB	400GB	800GB
Capacity - Compressed*	200GB	400GB	800GB	1600GB
Transfer Rate - Native	10-20MB/s	20-40MB/s	40-80MB/s	80-160MB/s
Transfer Rate - Compressed*	20-40MB/s	40-80MB/s	80-160MB/s	160-320MB/s
Media	Metal Particle	Metal Particle	Metal Particle	Thin Film

Figure 54. Ultrium migration path

The **LTO Accelis format** will utilize all the advantages of LTO described for Ultrium, except that the capacity has been reduced for the purpose of improving the access time to data dramatically. The capacity reduction is the result of making the tape narrower (hence fewer tracks) and by putting two reels in the cartridge instead of one. This results in 216 meters of 8mm wide tape, which contains 25GB of data (>50GB compressed). The dual reel cartridge implementation provides a number of advantages.

- The tape does not have to be threaded around to a take-up reel at all, enhancing the reliability of Accelis, which is particularly critical in a library environment.

- Because the tape path is fully contained in the cartridge, the tape can be loaded at the mid-point, which, in conjunction with the eliminated thread time, improves search times to sub-10 second averages.

Accelis records the LTO format in two data bands, instead of four as with Ultrium, and each data band contains 128 tracks (256 tracks total). Like Ultrium, the data bands can be recorded with eight-channel or four-channel heads and eight servo positions are defined in the servo band to support the recording.

The most unique aspect of Accelis cartridge design is that this media can be pre-positioned or post-rewound to further improve system performance. Also, in conjunction with the LTO-CM, an offline device (such as a smart picker or a queuing device) can locate to a specific point on tape, prior to tape drive insertion. This is a very powerful concept that will enable unprecedented solutions unique to Accelis. Figure 55 shows the expected migration path for Accelis.

	Generation 1	Generation 2	Generation 3	Generation 4
Capacity - Native	25GB	50GB	100GB	200GB
Capacity - Compressed*	50GB	100GB	200GB	400GB
Access Time	<10 seconds	<8 seconds	<7 seconds	<7 seconds
Transfer Rate - Native	10-20MB/s	20-40MB/s	40-80MB/s	80-160MB/s
Transfer Rate - Compressed*	20-40MB/s	40-80MB/s	80-160MB/s	160-320MB/s
Media	Metal Particle	Metal Particle	Metal Particle	Thin Film

Figure 55. Accelis migration path

Chapter 7. Disk storage subsystems

This chapter addresses the current disk storage solutions available for the most popular platforms in the market. Table 4 provides the connection capability of the various disk subsystems to the servers.

Table 4. Connection and reference table

Disk Subsystems	S/390	Open Systems	Reference
Enterprise Storage Server	Yes	Yes	7.1 on page 121
7133 Serial Storage Architecture	Yes(*)	Yes	7.2 on page 132
IBM Modular Storage Server	No	Yes	7.3 on page 135
IBM FAStT200 Storage Server	No	No(**)	7.4 on page 142
RAMAC Virtual Array Turbo Model X83	Yes	No	7.5 on page 144

* Only if attached through an Enterprise Storage Server

** Only for Intel based Servers, such as Netfinity.

7.1 The IBM Enterprise Storage Server (ESS)

The Enterprise Storage Server (ESS) is the latest IBM storage product to be developed using IBM's Seascape architecture (Figure 56).

The Seascape architecture is the key to the development of IBM's storage products both now and in the future. Seascape allows IBM to take the best of the technologies developed by the many IBM laboratories and integrate them to produce flexible and upgradable storage solutions — technologies such as the PowerPC, the Magstar Tape drives and IBM's award-winning Ultrastar disk drives. Serial interconnect technology, SSA, now delivering a 160 MB/sec data rate (see 6.2.2, "Serial Storage Architecture" on page 83), is used within the Seascape architecture to connect the processors and internal disks to ensure high performance and availability. Seascape also allows us to adapt to new technologies quickly, such as Fibre Channel and S/390 FICON.
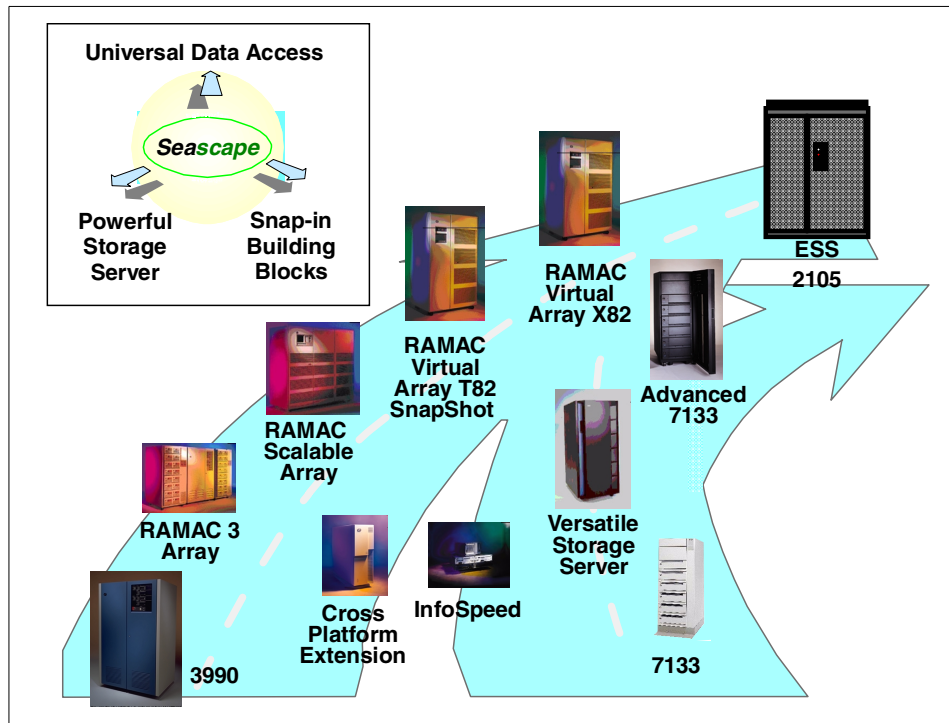
Figure 56. Positioning of the Enterprise Storage Server

As shown in Figure 56, the Enterprise Storage Server is the natural successor to the IBM 3990. It will provide all the functions that were available on the 3990, including peer-to-peer remote copy (PPRC), extended remote copy (XRC), and concurrent copy. For the many customers that have installed the RAMAC Virtual Array (RVA), with its revolutionary log structured file (LSF) architecture, we have protected their investment in this technology too.

The Enterprise Storage Server will, in the future, implement an LSF architecture. You will even have the choice of how much capacity you have under LSF arrays, and how much under the existing ESS design. If you have implemented SnapShot, then your investment in the function will be protected on ESS—initially by a function (FlashCopy) that will provide fast, real, time-zero (T0) copies; and later when we have LSF, by the ESS SnapShot equivalent. We have made the software interface transparent to allow you to exploit RVA SnapShot, ESS FlashCopy or even Concurrent Copy without changing your procedures.

The Enterprise Storage Server's high performance, attachment flexibility, and large capacity enable you to consolidate your data from different platforms onto a single high performance, high availability box. Storage consolidation can be the first step towards server consolidation, reducing the number of boxes you have to manage and allowing you the flexibility to add or assign capacity when and where it is needed.

Virtually all types of servers can concurrently attach to the Enterprise Storage Server-including S/390, Windows NT, Windows 2000, Novell NetWare, AS/400, and many types of UNIX servers. With a capacity of up to 11 TB, and up to 32 host connections, an ESS can meet both your high capacity requirements and your performance expectations. You should consider merging your Open Servers and S/390 storage management teams to get the best of both worlds servicing your core business. See 16.1, "Understanding enterprise storage administration" on page 305 for more information.

7.1.1 Overview

The IBM Enterprise Storage Server is a high performance, high availability, high capacity storage subsystem. It contains two 4-way RISC processors with up to 16 GB of cache and 384 MB of non-volatile storage to protect from data loss. The ESS has a maximum capacity of over 11 TB with the second frame attached. Connectivity to S/390 is through up to 32 ESCON channels; and to UNIX, AS/400, or NT hosts, through up to 32 SCSI interfaces; or a combination of the two. The ESS will support both S/390 Fibre Channel (FICON) and Fibre Channel Protocol (FCP), including Fibre Channel Arbitrated Loop (FC-AL) and FC-switched for open systems.

7.1.1.1 Enterprise Storage Server models

There are two current models of the IBM Enterprise Storage Server:

- The IBM 2105-F20 Enterprise Storage Server

  This model has two three-phase supplies and supports the full complement of 128 disks in two 2105 cages.

- The IBM 2105-F10 Enterprise Storage Server

  The IBM 2105 Model F10 has single phase power and 420 GB to 1.7 TB approximate end-user capacity, in addition to increased scalability, throughput, and performance. These occupy a single 2105 cage.

Shown in Figure 57 is a photograph of a 2105-F20 Enterprise Storage Server (ESS) with the covers removed.
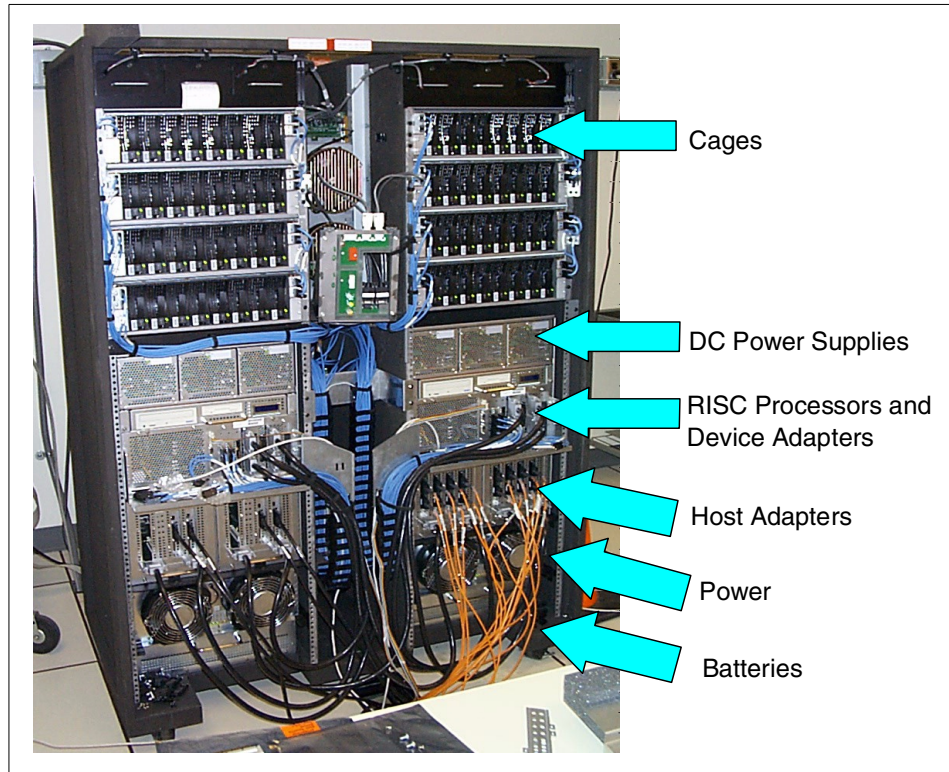
Figure 57. The IBM Enterprise Storage Server

At the top of the frame are the disks, below that the DC power supplies, and directly below these are the RISC processors. Just below the processors are the Device Adapters and the Host Adapters clearly shows the two clusters, one on each side of the frame.

The height of the ESS is 70.7 inches (1.793 meters), width is 54.4 inches (1.383 meters), and depth is 35 inches (0.89 meters), without its top cover. The ESS requires a raised floor environment.

The ESS expansion rack is the same size as the ESS.

7.1.1.2 ESS expansion capability
- ESS Expansion Rack feature 2100

  This rack attaches to the 2105-E20 only and uses three-phase power supplies. Up to four ESS cages can be installed in the 2105-E20 expansion rack. This gives a maximum of 256 disks in the four cages.

- IBM 2105-100 Versatile Storage Expansion Rack

  Up to three IBM 2105-100 VSS Expansion Racks can be attached to an Enterprise Storage Server. Each expansion rack supports up to seven 7133-D40 or 7133-020 Drawers. The drawers must contain full configurations of 16 disks and are attached to the ESS Device Adapters.

  Maximum capacity for any combination of 2105-F10, F20 or 2105-100 Expansion Racks is 384 disks.

7.1.1.3 Statement of general direction

IBM plans continuing growth and enhancements to the Enterprise Storage Server, and to make these enhancements available as upgrades to installed Enterprise Storage Servers:

- Native FICON support on S/390
- SDD enhancements
- Increased Storage Capacity
- Continued performance enhancements

This statement represents IBM's current intent and is subject to change or withdrawal.

7.1.2 Key features

The IBM Enterprise Storage Server is a superior SAN storage solution that:

- Supports critical requirements for strategic business initiatives, including:

  - e-business
  - Enterprise resource planning
  - Business intelligence
  - Server consolidation
  - All other mission-critical applications

- Deploys the IBM Seascape principles:

  - Powerful storage server
  - Snap-in building blocks
  - Universal data access

- Builds on the rich legacy of IBM storage systems solutions across the major operating system platforms

- Introduces new innovations to deliver substantial, unique business advantages

- Provides heterogeneous attachment to all major server platforms, including S/390, AS/400, Windows NT, and many varieties of UNIX

- Supports Fibre Channel attachment capability
- Has excellent scalability:
  - From 400 GB to over 11 TB
  - Simple selection from 16 standard configurations to meet capacity and performance needs
- Delivers superior performance that can be optimized to meet your changing heterogeneous environment needs:
  - High bandwidth and advanced transaction processing capabilities provide solutions for both online and batch applications
  - Innovations to reduce resource contention and dramatically improve performance:
    - Parallel Access Volumes
    - Multiple Allegiance
    - I/O Priority Queuing
- Supports 7 x 24 availability requirements with:
  - High-availability architecture
  - Redundant componentry
  - Extensive data replication services:
    - Access to data while making a copy using Concurrent Copy
    - Business continuity through remote copy services — PPRC and XRC
    - Rapid data duplication through FlashCopy, providing extensive capabilities to exploit, manage, and protect your information in a 7 x 24 environment
- Works in conjunction with the IBM StorWatch ESS Specialist and ESS Expert, comprehensive Web-based tools for:
  - Configuration
  - Performance
  - Asset and capacity management
  - Facilitation of storage area network operations
- Maximizes storage server availability through redundancy and nondisruptive service with design to virtually eliminate single points-of-failure or points-of-repair

See *IBM Enterprise Storage Server*, SG24-5465 for more information.

7.1.3 Specifications

Figure 58 shows a schematic of the Enterprise Storage Server.
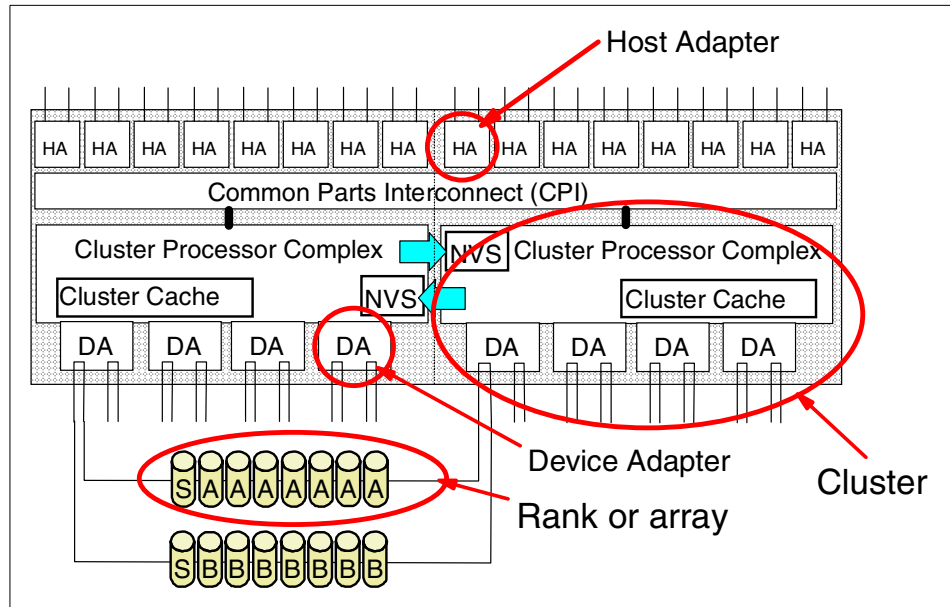
Figure 58. ESS schematic view

At the top we have up to 16 host adapters (HAs), each HA supporting two ports. Each HA is connected to both clusters through the Common Parts Interconnect (CPI) busses, so that either cluster can handle I/O from any host adapter. You can see the two clusters that contain the RISC processors, the cache, and the NVS for the opposite cluster. Within each cluster we have one to four device adapters (DAs). They are always installed in pairs, and the disk arrays (or *ranks*, as they are called in the ESS) are attached through an SSA loop to both DAs in a pair. The ranks can be configured as RAID arrays, or as Just a Bunch Of Disks (JBOD).

A Logical Subsystem (LSS) consists of one or more ranks on an SSA loop specified as either CKD or FBA.

7.1.3.1 Performance

The Enterprise Storage Server is designed as a high performance storage solution and takes advantage of IBM's leading technologies. In today's world, where your business can reach global markets through e-business, you need business solutions that can deliver high levels of performance continuously every day, day after day. You also need a solution that can handle different workloads simultaneously, so you can run your Business Intelligence models, your large databases for Enterprise Resource Planning (ERP), and your online and Internet transactions alongside each other with minimal impact. See *IBM Enterprise Storage Server*, SG24-5465 for more performance information.

7.1.3.2 Scalability

Snap-In Building Blocks solution: Choose one of 32 standard configurations to meet your business needs. Capacity of the standard configurations scales from 400 GB to over 11 TB:

- Two ultra-high performance configurations using 9 GB disk drives
- Fifteen configurations using 18 GB disk drives that deliver a variety of superior performance and bulk storage configurations
- Fifteen data "storehouse" configurations using 36 GB disk drives

Each standard configuration automatically provides the proper support infrastructure:

- 16 GB of cache
- 384 MB of nonvolatile storage (NVS)
- Four Device Adapter pairs
- Serial 9 GB, 18 GB, or 36 GB disks in the optimum serial loop configuration

Portions of the storage may be configured as RAID-5 protected or non-RAID (storage used by AS/400 must be RAID-5). Emulated LUNs are used for UNIX, Windows NT, and AS/400 server storage. System/390 storage may be identified as 3380 or 3390 track format volumes.

In addition, to provide investment protection for existing 7133 Serial Disk and Versatile Storage Server capacity (see A.2, "Versatile Storage Server" on page 367), the Enterprise Storage Server supports reuse of IBM 2105 Model B09 and 100 frames containing IBM 7133 Serial Disk Model 020 and D40 drawers (see 7.2, "7133 Serial Disk Subsystem Advanced Models D40 and T40" on page 132) in your Enterprise Storage Server configuration.

7.1.3.3 Connectivity

Currently, the Enterprise Storage Server supports FICON through the IBM 9032 Model 5 ESCON Director FICON Bridge card. This card supports up to 8 ESCON paths over a single Fibre Channel link.

As part of the Enterprise Storage Server announcement, IBM is previewing the support for Fibre Channel, including FICON, on the ESS (see 6.2.5, "Fiber Connection (FICON)" on page 96).

The Enterprise Storage Server provides an Ultra SCSI interface with the SCSI-3 protocol and command set for attachment to UNIX systems, Windows NT, and AS/400. This interface also supports SCSI-2.

Hosts supported by the ESS SCSI interfaces include:

- IBM RS/6000
- HP 9000 Series
- Sun
- IBM Netfinity
- Data General
- IBM AS/400

For a current list of supported SCSI Hosts, visit the IBM Web site at:

    http://www.storage.ibm.com/hardsoft/products/ess/supserver.htm

7.1.3.4 Manageability

The StorWatch ESS Specialist comes with the ESS product. The ESS Specialist must be used to configure an ESS. Apart from the configuration function for ESS, the ESS Specialist can be used to administer Copy Services functions to set up FlashCopy T0 copies or Peer-to-Peer Remote Copy (PPRC).

The StorWatch ESS Expert is an optional product. Its purpose is asset management, capacity management, and performance management for the ESS.

7.1.3.5 Availability

The Enterprise Storage Server has been designed with no single points of failure. It is a fault tolerant storage subsystem, which can be maintained and upgraded concurrently with customer operation. Figure 58 on page 127 shows the architecture of the ESS.

Beside the hardware part, ESS consolidates existing availability solutions such as remote copy, concurrent copy, and brings new ones such as FlashCopy that contribute a lot to customer data availability.

7.1.3.6 Copy Services

PPRC is part of Copy Services that is a separately sold feature of the ESS. It uses a synchronous protocol that allows real-time mirroring of data from Logical Unit (LUN) to another LUN. The protocol guarantees that the secondary copy is up to date by ensuring the primary copy is only written once it has received acknowledgment that the secondary copy has been written to. The LUNs can be in either the same ESS or in another ESS, hence the ability for a true remote copy to exist. PPRC is currently supported on many Unix platforms, Windows NT 4.0 and S/390.

FlashCopy is also a part of Copy Services. It makes a single point in time (T0)copy of a LUN. The process takes only a few seconds and becomes instantly available for use. It is generally used for on-line backups, creating a database for data-mining purposes or new application testing. FlashCopy is available for AIX, Solaris, HP-UX, Windows NT 4.0 and S/390 environments.See *Implementing ESS Copy Services on UNIX and WIndows NT*/2000, SG24-5757 for more information

Extended Remote Copy (XRC) provides a copy of OS/390 data at a remote location via telecommunications lines. Being asynchronous, XRC is not linked by any distance. For a detailed explanation of XRC, and how it is implemented, please refer to "Chapter 3. XRC", *Implementing ESS Copy Services on S/390,SG24-5680.*

Concurrent Copy is a copy function for the OS/390 operating system, that enables data center operations staff to take a copy or a dump of data while applications are updating that data. Similar to FlashCopy, it creates a T0 copy of the source. Concurrent Copy, however, can act not only on a full volume, but also on data sets. For more information on Concurrent Copy, please refer to "Chapter 5. Concurrent Copy", *Implementing ESS Copy Services on S/390,SG24-5680*

7.1.3.7 Performance Enhancement Features for S/390

ESS offers the ESS EX Performance Package—this consists of three performance features:

- Multiple Allegiance
- Parallel Access Volumes (PAV), which is a chargeable feature
- I/O Priority Queueing

Multiple Allegiance — The Multiple Allegiance feature allows the ESS to accept an input/output (I/O) request from each channel path group for a unit address. The request can be concurrent with other I/O requests. The ESS queues requests and then initiates them. This allows you to have multiple access to a shared volume (logical device) without affecting performance.

Using the Multiple Allegiance feature can improve ESS throughput and performance.

See the redbook, *IBM Enterprise Storage Server, SG24-5465* for more information on Multiple Allegiance.

Parallel Access Volumes (PAV) —The Parallel Access Volume feature allows you to access a CKD volume on an ESS in parallel. With PAV, you assign a base address to a CKD logical device. You can use any of the 256 available addresses as the base device address. You then configure alias addresses for the CKD logical device.

You can then dynamically reassign any configured alias address from one CKD logical device to another CKD logical device. Address types are:

 • Base address

The base address is the unique unit address of a volume. The ESS associates storage space with this address.

 • Alias address

You can assign several alias addresses to the base address. The ESS runs I/O operations to the alias address against the base address.

With PAV, a single System/390 can simultaneously process multiple I/O operations to the same logical volume. This can reduce volume contention.

See the *IBM Enterprise Storage Server User's Guide* for more information on using PAV.

I/O Priority Queueing — I/Os from different system images can be queued in a priority order. Again it is OS/390's Workload Manager that can utilize this priority to favor I/Os from one system against the others. You can activate I/O Priority Queuing in WLM's Service Definition settings (see Figure 90 on page 121). Obviously, the WLM has to run in Goal mode. All Releases of OS/390 Version 2, and OS/390 Version 1, Release 3, support I/O Priority Queuing.

See the redbook, *Enterprise Storage Server, SG24-5465* for more information on I/O Priority Queuing.

7.2 7133 Serial Disk Subsystem Advanced Models D40 and T40

The IBM 7133 Serial disk subsystem is an advanced storage subsystem capable of connecting and sharing storage with multiple UNIX and Windows NT hosts.

Through its implementation of industry-standard serial technology, the 7133 Advanced Models provide outstanding performance, availability, multihost attachment, and economical RAID 5 data-protected storage.They are also building blocks in the IBM Seascape architecture, a blueprint for innovative storage solutions.

7.2.1 Overview

The 7133 Advanced Models subsystem can contain four to sixteen drives per 7133 and one to six 7133 enclosures per 19 inch rack. Each 7133 has a Storage Serial Architecture (SSA) four-port adapter that allows two loops and delivers up to 40 MB/sec per adapter (thus, up to 160 MB/sec per subsystem). See Figure 59.
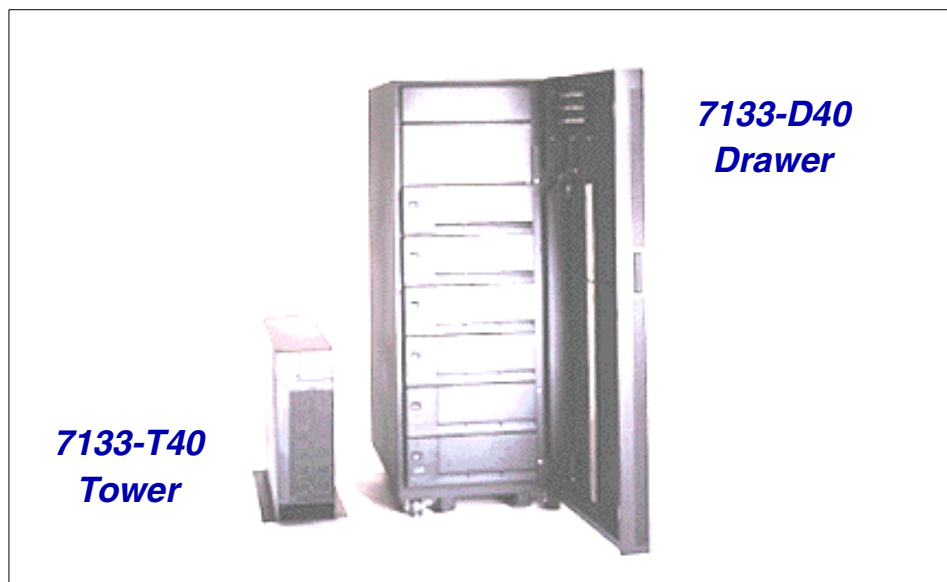
Figure 59. 7133 Serial Advanced Models D40 and T40

The 7133 Advanced Models subsystem can also attach up to 96 SSA disk drives, 48 per loop and can be configured with up to 8 adapters in up to 8 host processors within each SSA loop. The SSA RAID adapters can be configured with multiple RAID 5 arrays.

7.2.2 Key features

The 7133 Advanced Models subsystem has the following key features:

- SSA multihost attachments.

  By using the appropriate SSA adapters or the IBM 7190 SSA loop attachment, multiplatform servers can benefit from SSA technology. Refer to `www.ibm.com/storage/7133attach` for a current listing of supported product models.

- High performance, up to 80 MB/sec per loop.

  The 7133 Advanced Models provide superior performance because they offer greater bandwidth for multiple, concurrent, full-duplex I/O operations. Unlike SCSI bus configurations, SSA devices are configured in loops and do not require bus arbitration. This enables multiple concurrent operations to occur in separate sections of the loop, resulting in higher overall throughput.

- High capacity IBM Ultrastar disk drives providing up to 582 GB (36.4x16) per tower or drawer and (36.4x16x6) 3.5TB per host adapter.

- Fault tolerance with dual full-duplex ports, redundant paths, power and cooling, and hot swappable disk modules without quiescing I/O on any other disk drive module within the same loop. To further increase availability, the 7133 Advanced Models monitor and provide detailed information on the status of power, cooling, and disk drives.

- Easier management with StorWatch Serial Storage Expert.

  StorWatch Serial Storage Expert can also be used very effectively with the 7133 Advanced Models. The Serial Storage Expert will be able to automatically discover and provide information concerning the configurations of the attached disk storage. This information includes which disks are connected to which adapter and where those disks are located in the drawer or tower. This automatic discovery of detailed drive location information is only available with the D40 or T40 Advanced Models of the 7133 when used with the StorWatch Serial Storage Expert. To know more about StorWatch, see Chapter 14, "StorWatch" on page 287 for detail.

7.2.3 Specifications

SSA is a serial link with a two-signal connection (transmit and receive) providing full duplex communication.

There are significant availability benefits inherent in the SSA technology that enhance the IBM 7133 solution. With the SSA loop design, a single cable failure will not necessarily cause loss of access to data. If there is a failure on a loop, the SSA adapter automatically reconfigures itself and continues accessing the devices without the loop configuration. After the path is restored, the adapter automatically reconfigures to resume normal operation. If there is a disk failure, the hot swappable disks can be replaced without the loss of communication between the adapter and the other disks on the loop. The 7133 is also designed with redundant power supplies and fans that can be replaced without degrading system operations.

The rack-mountable 7133 Advanced Model D40 is designed for integration into an industry standard 19 inch rack, while the 7133 Advanced Model T40 is a free-standing deskside tower unit (see Figure 59 on page 132).

The 7133 Advanced Models D40 and T40 can be populated with 4.5 GB, 9.1 GB, 18.2 GB, and 36.4 GB Advanced SSA drive modules. Module capacities can be intermixed, enabling the flexibility to build storage environments from gigabytes to terabytes. Each 7133 includes three auto-docking power/cooling units, which provide redundant power to all sixteen disk drive module positions.

An Advanced SSA Optical Extender (Feature 8851) is available on all Advanced Models. This feature supplies a pair of optical converters that convert electrical SSA signals into optical signals for transmission using optical fibers. This feature allows the maximum distance between SSA nodes to be increased from 25 m up to 10km.

In addition, to provide investment protection for existing 7133 Serial Disk and Versatile Storage Server capacity, the Enterprise Storage Server supports reuse of IBM 2105 Model B09 and 100 frames containing IBM 7133 Serial Disk Model 020 and D40 drawers in your Enterprise Storage Server configuration (see 7.1, "The IBM Enterprise Storage Server (ESS)" on page 121). Designed to facilitate storage resource sharing, the Enterprise Storage Server provides a way to consolidate disk storage into a single powerful system, delivering many levels of advanced functionality such as remote Web-based management and true data sharing for like servers.

7.3 IBM Modular Storage Server

In this section we present a high-level overview of the IBM Modular Storage Server (MSS) and its supporting software. For more information, see *Introduction to the IBM Modular Storage Server*, SG24-6103.

The MSS consists of dual controllers, redundant power supplies, 4 TB storage capacity, up to 1 GB of cache, and heterogeneous Fibre Channel support (see Figure 60).

Figure 60. IBM Modular Storage Server

7.3.1 Overview

The Modular Storage Server is a fibre attached Raid controller. It utilizes six SCSI buses on the backend to attach hard drive storage into a fibre channel network. Supported Raid levels include Raid 0, Raid 1, Raid 3/5(automatically uses synchronous data transfer from the Raid 3 specification when applicable), and Raid 1+0.

Advanced features include FlashCopy, Peer-to-Peer Remote Copy (PPRC), and device cloning (copying a given LUN and breaking it as an independent device).

The system is similar to the Compaq MA8000. While some of the components are basically the same they are not the same offering, IBM disk drives are used extensively in the MSS. Keep this in mind if you also reference Compaq's documentation on the MA8000.

7.3.1.1 Architecture

Figure 61 shows the design of the MSS in schematic form.
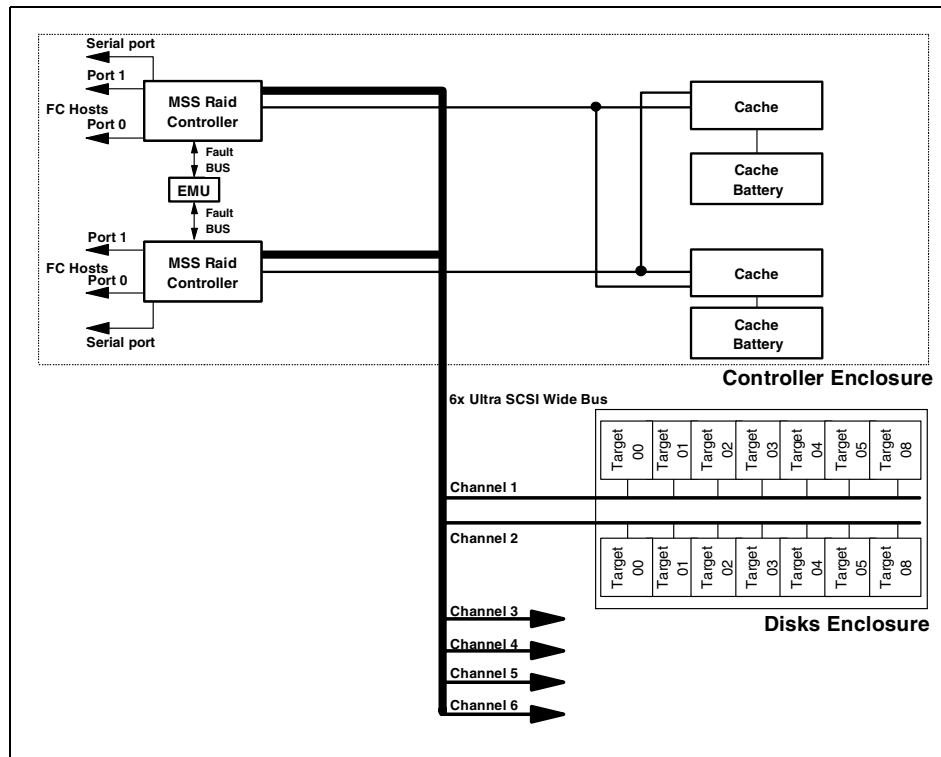
Figure 61. IIBM Modular Storage Server Architecture

7.3.2 Key features

The MSS provides the following interesting features:

- Multihost attachment (see 7.3.2.1)
- High-availability design (see 7.3.2.2)
- Extended distances (see 7.3.2.3)
- High performance (see 7.3.2.4)
- Dynamic management (see 7.3.2.5)
- Expandability (see 7.3.2.6)
- Data replication (see 7.3.2.7)

7.3.2.1 Multihost attachment

The IBM Modular Storage Server is a flexible, scalable, and easy-to-manage storage server for Fibre Channel Storage Area Networks (SANs) Provides multiplatform support for major UNIX, Windows NT, Windows 2000, Novell NetWare, and OpenVMS platforms in homogeneous or heterogeneous multihost environments. The Modular Storage Server is ideal for midrange servers running critical applications that require a storage system with built-in redundancy and fault-tolerance. Please reference the following Compaq documentation for details in planning and implementing heterogeneous connectivity. *Compaq StorageWorks Application Note Heterogeneous Storage Area Networks, EK-SMA30-AN.xxx*

7.3.2.2 High-availability design

As access to information becomes more critical than ever, companies are demanding flexible, high-performance, continuously available storage systems. With RAID levels RAID levels 0, 1, 0+1, and 3/5 to choose from, the IBM Modular Storage Server can help you reach the optimum system performance and protection for your business-critical applications. Its dual-active RAID controllers provide high throughput and redundancy. Utilizing the MSS Data Path Optimizer (DPO), increases data availability through the use of redundant paths To increase availability, each RAID controller supports up to 512 MB of battery-backed cache. Dual fans and power supplies further contribute to its ability to support 24 x 7 operations. Components such as controllers, fans, drives, and power supplies are hot pluggable, simplifying access and maintenance for continuous operations.

7.3.2.3 Extended distances

Industry-standard Fibre Channel host attachment provides the flexibility to configure the Modular Storage Server at distances of up to 10 km from the host server, which is suitable for disaster recovery implementations.

7.3.2.4 High performance

High-speed Fibre Channel attachment meets the requirements of high-performance networked systems running large database applications, such as enterprise resource planning and data warehousing. The IBM Modular Storage Server offers high performance with four 100 MB/sec fibre interfaces and dual active controllers. The performance is also enhanced when combining it with the optional 1 GB of internal cache and utilizing 18 GB, 15,000 RPM disk drive modules.

7.3.2.5 Dynamic management

The StorWatch MSS Specialist is a network-based integrated storage management tool that lets storage administrators configure, monitor, dynamically change, and manage multiple Modular Storage Servers from a single Windows2000 or WindowsNT workstation (see Chapter 14, "StorWatch" on page 287). High availability and full redundancy are provided with the host-specific Fibre Channel Storage Manager tool, which resides on the host system and provides automatic I/O path failover when a host adapter, IBM Fibre Channel Storage Hub, or a storage server controller fails.

7.3.2.6 Expandability

The IBM Modular Storage Server can expand up to 72 disk drives by placing a maximum of six Disk Drive Enclosures. It features in capacities of 18 GB, 36 GB, or 72GB. Storage capacity can start at 72 GB and grow to over 4 TB.

7.3.2.7 Data replication

The MSS provides a robust portfolio of copy solutions. MSS *Peer-to-Peer Remote Copy* (PPRC) copies data in real time to a remote MSS that can be located thousands of miles from the primary site for disaster recovery.

MSS *FlashCopy* provides a user-friendly interface for point-in-time copying tasks with minimal impact on the production application. It manages controller-based cloning and instant copying operations to help minimize the downtime required to perform system backups and data migration activities.

ImageCopy is available by utilizing the CLONE utility. *ImageCopy* creates a physical copy of a storage unit to a new one and is available with the base code levels.

For more information please refer to "Chapter 7. Copy Functions" in the redbook *Introduction to the IBM Modular Storage Server, SG24-6103.*

7.3.3 Specifications

This section addresses:

- IBM Modular storage solution (see 7.3.3.1)
- Product characteristics (see 7.3.3.2)
- Software components (see 7.3.3.3)
- Supported platforms (see 7.3.3.4)

7.3.3.1 IBM Modular storage solution

The components of the IBM Modular storage solution include the following:

- 2106-200 IBM MSS Controller Enclosure — Four Fibre Channel ports for attachment to host servers. Two Fibre Channel ports per Raid Controller. Dual active RAID controllers support up to 512 MB of battery-backed cache. RAID levels of 0, 1, 0+1, and 3/5 are supported. Redundant power and cooling are standard.

- 2106-D10 IBM MSS 10 Bay Single Bus Disk Drive — Provides 10 slots for 1.6-inch disk drive modules. The D10 model supports up to 720 GB using the 72 GB/10K rpm Disk Modules. The D10 model contains a single SCSI bus and single SCSI port for connecting to the MSS Controller Enclosure

- 2106-D14 IBM MSS 14 Bay Single Bus Disk Drive — Provides fourteen slots for 1.0-inch disk drive modules. The D14 model contains a single SCSI bus and single SCSI port for connecting to the MSS Controller Enclosure.

- 2106-D20 IBM MSS 10 Bay Dual Bus Disk Drive — Provides 10 slots for 1.6-inch disk drive modules. The D20 model contains dual SCSI buses and two SCSI ports for connecting to the MSS Controller Enclosure. The D20 model supports up to 720 GB using the 72 GB/10K rpm Disk Modules.

- 2106-D24 IBM MSS 14 Dual Bus Disk Drive Enclosure — Provides fourteen slots for a 1.0-inch disk drive modules. The D24 model contains dual SCSI buses and two SCSI ports for connecting to the MSS Controller Enclosure. A fully populated D24 model can support up to 504 GB using the 36 GB/10 K rpm Disk Modules.

7.3.3.2 Product characteristics
Product characteristics include the following:

- Storage capacity 72 GB minimum, 4 TB maximum

- Industry-standard Fibre Channel host attachment enables configuration as far as 10km

- Broad range data protection options with RAID 0, 1, 0+1, 3/5

- High performance with an aggregate bandwidth of 400 MB/sec

- Storage management with StorWatch MSS Specialist

- High availability through automatic I/O path failover

- Concurrent maintenance with hot-pluggable components:
    - Controllers
    - Fans
    - Drives
    - Power supplies

- Up to 1 GB of battery-backed dynamic cache.

7.3.3.3 Software components

The Modular Storage Server provides several software options. They consist of the following:

- StorWatch MSS Specialist (required, one per platform)
- MSS Control Module (required)
- MSS Control Module for FlashCopy (optional)
- MSS Control Module for Peer-to-Peer Remote Copy/FlashCopy (optional)
- MSS Data Path Optimizer (DPO) (optional)
- MSS Peer-to-Peer Remote Copy Manager (optional)
- MSS FlashCopy Manager (optional)

For more information on these software components please refer to the redbook *Introduction to the IBM Modular Storage Server, SG24-6103*.

7.3.3.4 Supported platforms

The MSS supports these IBM and non-IBM systems:

- Intel-based servers running Windows NT Server 4.0, Windows 2000, and Novell Netware 4.2 or 5.0
- Sun Microsystems SPARCserver and Enterprise servers with Sun Solaris 2.6 and 7.
- Hewlett Packard 9000 K Series server with HP-UX V10.20, 11.0
- Silicon Graphics Origin 2000 Server with IRIX 6.5.3
- Compaq Alpha Servers with Open VMS 7.2.1, Tru64 Unix 4.0F and 5.0A
- RS/6000 with AIX 4.3.3

Details on models and operating system releases are available on the Web at URL:

```
http://www.ibm.com/storage/mss
```

7.4 IBM FAStT200 Storage Server

If you need a low-cost, space-efficient storage solution for the exploding Internet markets, then you will be interested in the IBM FAStT200 Storage Server. This attractive disk storage server is ideal for small to medium businesses that require a low-cost, full-fibre solution that is compact and Storage Area Network (SAN)-ready for Intel-based servers. You will also see this unit referred to as the IBM 3542 storage server.

The FaStT200 is illustrated in Figure 49.

Figure 62. IBM FaStT200 Storage Server (3542)

7.4.1 Overview

The IBM FAStT200 Storage Server provides flexible, affordable storage for Intel-based servers. The FAStT200 is designed for workgroup and departmental servers that require an external storage solution. The single controller model provides a cost-effective solution, while the FAStT200 High Availability (HA) model features a fully redundant configuration with dual-active controllers. As your storage requirements grow, you can easily expand storage capacity by adding IBM FAStT EXP500 Expansion Units.

7.4.2 Key features

FAStT200 provides a cost-effective, fibre-to-fibre, SAN-ready solution for Intel-based servers. Key features include:

- Four compact models to suit customer needs
- High-availability technology

- High-performance 1 Gb/s Fibre Channel technology

- Scalability

- Data storage up to 10 km (6.2 miles) away - provides additional protection from catastrophic occurrences

- Support for long-wave and short-wave Fibre Channel environments

- Supports up to 734 GB storage capacity per unit utilizing 10 K RPM 72 GB drives

- 15 K RPM 18.2 GB Fibre Channel drive option

7.4.2.1 Scalability

The FAStT200 Storage Server is available in two versions: Single controller FAStT200 (Models 1RX or 1RU) or dual controller, FAStT200 HA (Models 2RX or 2RU). The single controller version will provide low-entry price point, whereas the dual controller version should be implemented for high-availability needs. All configurations deliver up to 734 GB of storage capacity in a single 3U-high, rack-mount package.

The FAStT200 Storage Server supports up to two IBM FAStT EXP500 Expansion Units. Each EXP500 Expansion unit houses up to 10 fibre channel hot-swap disk drives to provide a storage capacity of up to 734 GB per unit. It supports 10,000 rpm 18, 36 or 72 GB FC HDDs, or the 15,000 RPM 18 GB FC HDD. The EXP500 is a fully-redundant, highly available unit with hot-swap dual power supplies and fans.

As a rack-mount solution, the FAStT200 can support up to ten Fibre Channel hard disk drives internally in a 3U (EIA units) space. It supports 18, 36, and 73 GB 10,000 RPM drives and 18 GB 15,000 RPM drives. The attachment of two FAStT EXP500 Expansion Units (containing up to 20 drives) increases storage capacity to more than 2 TB.

7.4.2.2 High availability

The FAStT200 helps ensure high availability by utilizing redundant, hot-swappable components--such as power supplies and fans. The FAStT200 HA model features dual-active RAID controllers with transparent failover support to further increase availability. With multiple RAID levels (0, 1, 3, 5, and 10), the FAStT200 can help protect valuable data and keep business-critical applications up and running. Also note that each RAID controller comes standard with 128MB of batter backed up cache.

With the FAStT200, you can maximize the availability of your servers by using Microsoft Cluster Service (MCS) and Netfinity Availability Extensions for MCS. For supported MCS configurations, visit

`http://www.storage.ibm.com/hardsoft/products/fast200/fast200.htm`

7.4.2.3 Extended distances

Industry-standard Fibre Channel host attachment provides the flexibility to configure the FaSt200 at distances of up to 10 km from the host server, which is suitable for disaster recovery implementations

7.4.2.4 Storage management

The IBM FAStT Storage Manager is a network-based, integrated storage management tool that enables administrators to configure, monitor, dynamically change, and manage multiple FAStT200 Storage Servers from a single Windows 95 or Windows NT workstation. The software provides remote management capabilities and the ability to share storage resources among servers through storage partitioning. The FAStT Storage Manager supports up to 32 logical units to increase configuration flexibility for large storage arrays. The FAStT200 supports up to four host systems or two two-node clusters.

7.4.3 Specifications

The specifications for the FaStT200 are as follows:

- Dual Raid Controllers with 128MB battery backed cache
- Dual hot pluggable power supplies and fans
- Fiber Channel attach for FC-AL, switched fabric or point-to-point,
- Raid level support of 0,1, 3, 5, 10
- Disk drives of 18.2GB, 36.4GB, and 73.4GB sizes supported

For more information and specification on the FaStT200, visit:

`http://www.storage.ibm.com/hardsoft/products/fast200/fast200.htm`

7.5 RAMAC Virtual Array Turbo Model X83

RAMAC Virtual Array Turbo Model X83 (RVA Turbo X83) is the latest device of the RVA family. As a new member of the RVA family, RVA Turbo X83 offers investment protection for your virtual architecture environment by also taking advantage of the advanced virtual disk architecture, providing high availability and performance, fault tolerance, and advanced functions, all at an affordable cost.

The RVA Turbo X83 is illustrated in Figure 63.

Figure 63. RAMAC Virtual Array Turbo Model X83

7.5.1 Overview

RVA Turbo X83 is built on the virtual disk architecture, and emulates both IBM 3380 J/E/K and 3390 1/2/3/9 disks. RVA Turbo X83 employs an advanced storage subsystem architecture where logical volumes are indirectly mapped to physical volumes within the subsystem. Unlike traditional storage subsystems, RVA Turbo X83 maps user data across all the disks in the subsystem.

Operating in this manner, RVA Turbo X83 is able to deliver outboard storage management and data reproduction functions not available in any other System 390 subsystem. In addition, it can more effectively use storage capacity by storing only data that is actually used and not reserving capacity for unallocated space or space that has been allocated, but is unused.

RVA Turbo X83 delivers better performance with faster microprocessors in the subsystem, larger cache size (up to 6GB), and with IBM Serial Disk Drives state of the art disk drives manufactured by IBM under the Seascape umbrella. Eight concurrent data transfers over ESCON channels,

turbo-shared memory, and new microcode are also strong contributors to the performance enhancement.

See the *IBM RAMAC 9393 General Information Manual*, GC26-7167 for more details.

A hardware diagram of the RVA Turbo X83 is shown in Figure 64.
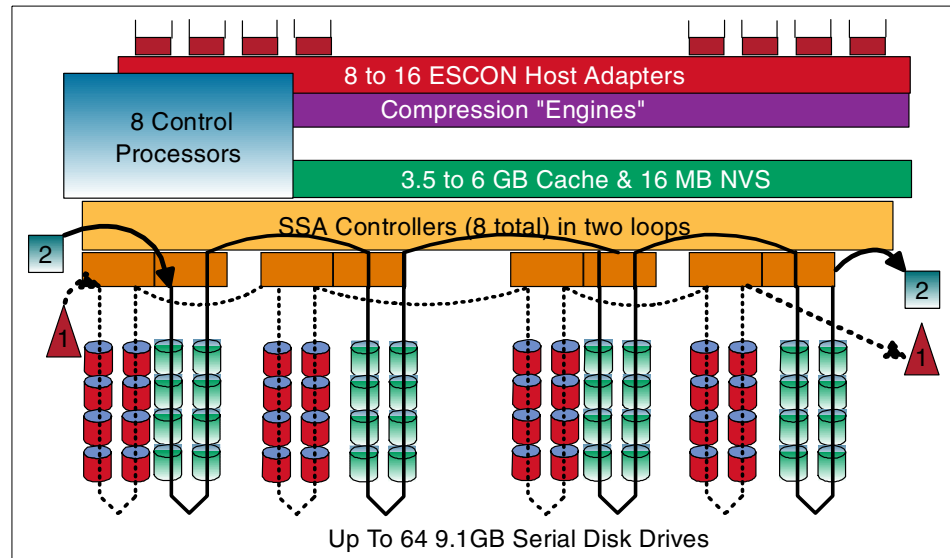
Figure 64. RVA Turbo X83 logical internal structure

7.5.2 Key features

The RAMAC Virtual Array Turbo Model X83 includes the following unique and useful features:

- 7.5.2.1, "Data compression" on page 148
- 7.5.2.2, "RAID 6" on page 148
- 7.5.2.3, "SnapShot" on page 149
- 7.5.2.4, "Peer-to-Peer Remote Copy support" on page 151

Over the last 30 years, disk storage architecture has mapped a logical disk volume to a real physical volume. To update a record, the record is read into storage, changed, and then written back to the exact same location on the physical disk.

RVA Turbo X83 uses virtual disk architecture. RVA Turbo X83 employs an advanced storage subsystem architecture that provides up to 1024 logical volumes, four times more than the previous subsystem, independent of subsystem storage capacity. Logical volumes are indirectly mapped to physical volumes within the subsystem. For example, on a write operation the control unit compresses the data and writes it to a new location on the next available disk. RVA Turbo X83 uses storage capacity more efficiently by storing only data that is actually used — not reserving capacity for unallocated space, or unused allocated space, or interrecord gaps. This unique way of storing data is called virtual disk architecture.

The *Log Structured File* (LSF) system is an important function for RVA's virtual disk architecture. LSF is the way to write data to a new location on an array at all times, making it possible to avoid the write penalty.

RVA Turbo X83 automatically spreads the workload evenly over all disks, virtually eliminating disk hot spots and dramatically reducing the need for traditional performance management.

You can define and delete logical volumes on RVA Turbo X83 as necessary while the subsystem is in use. With RVA's capability to define 1024 free logical volumes, you can reduce out-of-space conditions that result in application outages and minimize performance degradation resulting from logical volume I/O contention.

Even if the new capacity is brought up-to 1.68 TB, it is important to monitor the net capacity load (NCL), which is the ratio of physical capacity used to the entire disk capacity in RVA Turbo X83. We recommend keeping NCL less than 75%.

IBM eXtended Facility Product (IXFP) is host software that extends the advantage of RVA's virtual disk architecture. You can easily manage the RVA's configuration, such as logical volumes or disk arrays, with IXFP. It enables you to monitor NCL from host systems. IXFP also offers functions that report subsystem utilization. IXFP runs under MVS or VM. IXFP is required for the SnapShot function. IBM strongly recommends that you install IXFP for the purpose of efficient management of RVA.

The virtual disk architecture is illustrated in Figure 65.
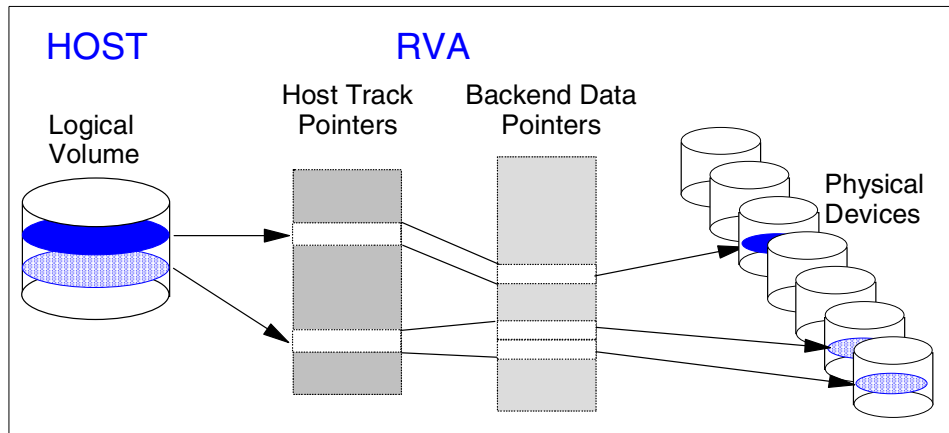
Figure 65. Virtual Array Architecture

7.5.2.1 Data compression

RVA Turbo X83 employs its own compression algorithm. RVA Turbo X83 compresses all write data at the channel interface using a modified Lempel-Ziv algorithm. Since all data stored in cache is compressed, data transfer rates among cache and disk arrays appear to be better than the original specification. In addition, RVA Turbo X83 eliminates the interrecord gaps associated with storing data on CKD or ECKD formats. Since these gaps are not stored on disk arrays, RVA Turbo X83 can compact data onto the disk arrays.

Operation of RVA's data compression process is transparent to the host. Since RVA Turbo X83 uses this sophisticated compression and compaction mechanisms to store data, storage capacities are quoted in *effective* or *usable* capacities. Actually realized storage will vary, depending on the actual data being compressed; some data sets will achieve higher ratios and some will achieve lower ratios. Users can often achieve compression ratios as high as 2:1 for cache/NVS and 3.6:1 for disk storage.

See 11.2.10, "Compression" on page 220 for details of data compression techniques.

7.5.2.2 RAID 6

RAID 6 architecture provides dual disk-drive redundancy, which enables it to continue providing user access to data even with two simultaneous disk failures in the same array. RVA automatically begins dynamic sparing at the array disk failure.

The effective array capacity of RVA Turbo X83 ranges from 290 GB to about 1.68 TB, depending on data characteristics. In addition, one RVA Turbo X83 subsystem includes two global spare disks. RVA Turbo X83 employs IBM 9.1 GB serial disk drives.

The comprehensive fault-tolerant design of RVA Turbo X83 helps retain access to data and system performance in the event of a hardware failure. Use of RAID 6, hot pluggable drives, power supplies, and system components practically eliminates unscheduled subsystem outages.

7.5.2.3 SnapShot

RVA Turbo X83 enables almost instantaneous views of data through SnapShot high-speed duplication. Unlike traditional solutions which can take minutes or hours to make a copy of data, SnapShot can create additional views of data in seconds.

SnapShot is an instantaneous copy function of RAMAC Virtual Array. RVA's virtual disk architecture enables you to produce almost instantaneous copies of data sets, volumes, and VM minidisks. We call this function *SnapShot*. SnapShot produces copies without data movement. We call making a copy with SnapShot a *snap*; the result of SnapShot is also called a *snap*.

SnapShot is a virtual data duplicator because it creates multiple views of data. Your server can see and access the original and the snap copies separately. SnapShot improves productivity within organizations by providing multiple views of data almost immediately. SnapShot-copied data can be used to enhance critical business applications, data mining, disaster recovery, and Year 2000 application development and testing.

Conventional methods of copying data on DASD consist of making a physical copy of the data on either DASD or tape. Host processors, channels, tape, and DASD controllers are involved in these conventional copy processes. Copying may take a lot of time, depending on available system resources.

With SnapShot, there is no movement of the actual data. Snapping can take just seconds rather than minutes or hours, because data is not moved and the host processor and the channels are not involved in the data transfer. Furthermore, additional physical disk space is not required to accommodate the snap until either the source or the target is updated. As far as the operating system is concerned, the snap is a real copy of the data; as far as the RVA hardware is concerned, it is a virtual copy of the data.

SnapShot is especially useful for Year 2000 testing. You can create the test data and backups with minimal additional disk space and a few seconds. SnapShot can accelerate the testing process.

SnapShot is available with IBM Extended Facilities Product (IXFP), SnapShot hardware feature, and SnapShot software. SnapShot can be used on both OS/390 environments and VM/ESA environments.

SnapShot at volume level is illustrated in Figure 66.
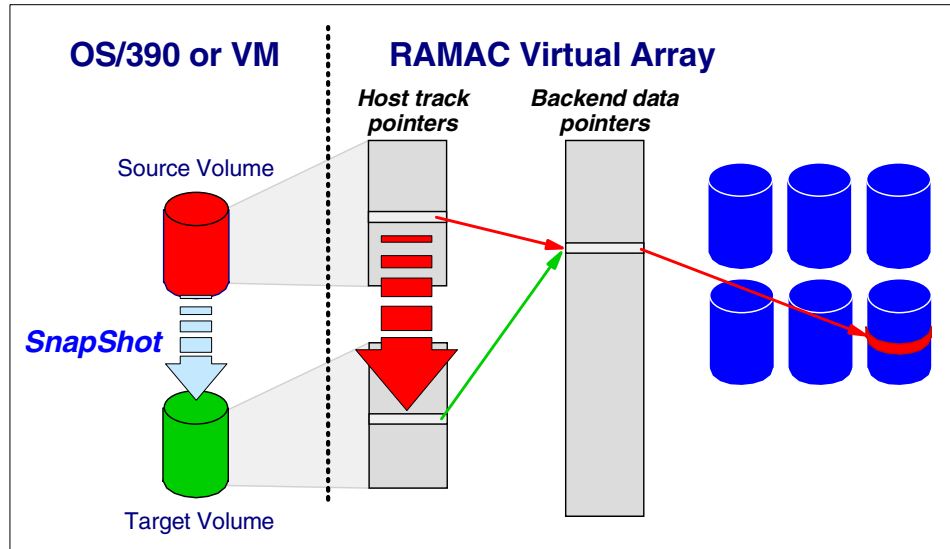
Figure 66. SnapShot at volume level

SnapShot data types
You can snap at both a volume and data set level in an OS/390 environment or at minidisk level in a VM environment. Volume snap can be performed from any source volume within the RVA to a target volume, which must be of the same device type and model and reside in the same RVA. Volume snap copies all data set types on the volume. When snapping a data set, the source data set and the target data set must reside in the same RVA. The target must be on the same device type as the source data set, but it can be on a different model.

Data set snap can be performed on the following data set types:

- Sequential and extended format sequential data sets
- Partitioned data sets and PDSEs
- Direct access data sets
- Multivolume data sets
- VSAM data sets, ESDSs, KSDSs, LDSs, RRDSs, and VRRDSs.

See *Implementing SnapShot*, SG24-2241 for more details.

Concurrent Copy-compatible SnapShot
You can invoke the SnapShot function through DFSMSdss concurrent copy command. We call this function CC-compatible SnapShot. This is useful for those who currently use Concurrent Copy in order to develop the image copies of databases.

Support storage subsystem
RVA Turbo X83 with SnapShot hardware feature is required for the SnapShot function.

7.5.2.4 Peer-to-Peer Remote Copy support
With the Peer-to-Peer Remote Copy (PPRC) functions, a copy of the data located at a remote site is kept in synch with the data at the primary site. If one site is lost in a disaster, data at the other site is available to rapidly recover critical business applications. PPRC is implemented in control unit hardware, so it is transparent to all host software.

See "Peer-to-Peer Remote Copy" on page 376, for details about PPRC.

7.5.3 Specifications
The specifications of RVA Turbo X83 are as follows:

- Effective storage capacity (min/max): 290GB/1.68 TB
- Effective cache capacity (min/max):3.5/6 GB
- Effective nonvolatile (NVS) capacity: 16 MB
- ESCON channels (min/max/logical paths): 8/16/128
- Logical volumes (max): 1024
- Device emulation: 3380 Models J/E/K, 3390 Models 1/2/3/9

Chapter 8. Removable storage solutions

This chapter addresses the current removable storage solutions available for the most popular platforms in the market. Table 5 provides, at a glance, the connection capability of the various removable subsystems to the servers.

Table 5. Connection and reference table

Tape Subsystems	S/390	Open Systems	Reference
3590 High Performance Tape Subsystem	Yes	Yes	8.1 on page 153
3490E Model F Tape Subsystem	No	Yes	8.2 on page 157
3494 Tape Library Data Server	Yes	Yes	8.3 on page 160
Virtual Tape Server	Yes	Yes	8.4 on page 166
3570 Magstar MP Tape Subsystem	No	Yes	8.5 on page 172
Linear Tape-Open (LTO)	No	Yes	8.6 on page 175
3575 Magstar MP Tape Library	No	Yes	8.7 on page 178
3995 Optical Library Data Server	Yes	Yes	8.8 on page 181

8.1 3590 High Performance Tape Subsystem

The IBM Magstar 3590 High Performance Tape Subsystem offers high performance, high capacity, and wide platform connectivity.

The Magstar 3590 tape subsystem (rack-mounted) is illustrated in Figure 67.

Figure 67. Magstar 3590 High Performance Tape Subsystem

8.1.1 Overview

The Magstar 3590 is designed for high-performance tape applications, including:

- High-speed data save operations, where backup windows are critical, and large amounts of data are archived to tape.

- Large-scale automated tape environments, where performance and reliability are required. These are primarily half-inch tape drive environments.

- Large-scale mass data archive applications, where massive amounts of data need to be quickly saved to tape storage and later recalled.

The Magstar 3590 uses a bidirectional *longitudinal* serpentine recording technique, and a new-generation magneto resistive head that reads and writes 16 data tracks at a time to 1/2 inch media. The tape makes eight back-and-forth motions and has 256 tracks in total.

Magstar 3590 tape drives use only Magstar 3590 cartridges, which — with the embedded LZ1 data compression technique — provide up to 120 GB capacity when extended media support is used (with 3:1 compression).

There is a choice of two cartridges.

- The 3590 High Performance Cartridge Tape, which provides up to 20 GB uncompressed.

- The 3590 Extended High Performance Cartridge Tape, which provides up to 40 GB uncompressed.

See 11.2.10, "Compression" on page 220, for more detail about data compression techniques.

The advanced Magstar 3590 tape drive is designed with a predictive failure analysis that tracks hard and soft errors on each cartridge and determines when a cartridge needs replacement or a drive needs servicing. This enhanced reliability can improve operations and reduce job reruns.

The Magstar 3590 Tape Drives can be attached to a SCSI host interface or Fibre Channel interface of selected IBM and non-IBM midrange and open systems. The tape drives can also attach to ESCON and FICON channels on host systems through a Magstar 3590 controller.

The Magstar 3590 tape drive is the cornerstone in a variety of tape solutions from standalone (3590 A14 or standard racks) to automated (IBM 3494 automated tape library including Magstar Virtual Tape Server). As a reusable

storage asset, Magstar 3590 tape drives can now operate in *StorageTek Silos*, leveraging the 3590 investment even further.

In the ESCON host-attached silo-compatible subsystem, up to four Magstar 3590 tape drives attach to a StorageTek Silo and communicate to the host through a 3590 Model A50 control unit. Configuration flexibility is enhanced even further, since each Model A60 can have one or four ESCON adapters. This enables the flexibility of selecting the control unit to drive configuration. A single 3590-A60 control unit with a single ESCON adapter attached to four 3590 tape drives (1x4) is the most cost-effective configuration. Alternatively, two A60 control units, each with two dual ESCON adapters, can be attached to eight 3590 tape drives (two 1x4s), providing even higher performance and availability.

Magstar 3590 can attach to a wide range of platforms; Hewlett-Packard, Sun, HP/Convex, CRAY, Silicon Graphics, Intel-based systems running Windows NT, as well as to IBM AS/400, RS/6000, and S/390 systems.

See the *IBM 3590 Tape Subsystem Technical Guide*, SG24-2506, and the *IBM 3590 Tape Subsystem: Multiplatform Implementation*, SG24-2594, for more details.

8.1.2 Specifications

The Magstar 3590 tape subsystem consists of the following models:

Magstar 3590 E1A tape drive — This drive is designed to be used in a Magstar 3494 tape library data server, Magstar Virtual Tape Server, and Magstar 3590 Silo-compatible tape subsystem.

Magstar 3590 E1A tape drive head reads and writes 16 data tracks at a time in eight back-and-forth motions, and has 256 tracks in total. Cartridges have an uncompacted capacity of 20 GB, with data compression and extended media providing up to 120 GB.

The 3590E1A Models will utilize the 3590 data cartridges in two ways:

- Read the current 128-track cartridge created by 3590 B Models. And, it can read them at the higher data rate of 14 MB/sec. Data stored for archive or retention purposes can be retained and read by the new models, thus preserving a sizable investment in data cartridges.

- Read and write on those cartridges in a 256-track format. Utilizing the new higher capacity design, the E Models can extend the current 10 GB cartridge capacity to 20 GB. With compression, this capacity can reach upwards of 120 GB (at 3:1 compression or higher).

Magstar 3590 E11 tape drive — This drive is rack and frame-mounted. It incorporates a 10-cartridge Automatic Cartridge Facility (ACF) for high-capacity unattended operation. Up to four Model E11s can be installed in a rack.

Connectivity, performance, and cartridge capacity are the same in the Magstar 3590 E Models. See Magstar 3590 E1A for details.

Magstar 3590 B1A tape drive — This drive is designed to be used in a Magstar 3494 tape library data server, Magstar Virtual Tape Server, and Magstar 3590 Silo-compatible tape subsystem.

This model includes an integrated SCSI-3 controller with two ports that support a 16-bit, fast-and-wide, Ultra SCSI differential interface. The model B1A tape drive has a 9 MB/sec native data rate (with compression up to 33 MB/sec sustained data rate).

The Magstar 3590 B1A tape drive head reads and writes 16 data tracks at a time in four back-and-forth motions, and has 128 tracks in total. Cartridges have an uncompacted capacity of 10 GB, with data compression and extended media support providing up to 60 GB. Magstar 3590 B1A can be field updated to the new E Models.

Magstar 3590 B11 tape drive — This drive is rack and frame-mounted. It incorporates a 10-cartridge Automatic Cartridge Facility (ACF) for high-capacity unattended operation. Up to four Model B11s can be installed in a rack.

Connectivity, performance, and cartridge capacity are the same in the Magstar 3590 B Models. See Magstar 3590 B1A for details. Magstar 3590 B11 can be field updated to the new E Models.

Magstar 3590 A60 controller — The 3590 Model A60 is a controller designed for attaching 3590 drives to ESCON and FICON channels on host systems. The A60 can be flexibly configured with up to four dual-ported ESCON adapter cards, up to two FICON adapter cards, or an intermix of ESCON and FICON adapters.

The Magstar 3590 Model A60 can be placed in multiple 3590 storage solutions. The 3590 A14 Frame with an integrated 3590 Model A60 Controller provides a frame attachment of up to four Magstar 3590 Tape Drives. It can be installed in a standard 19-inch rack such as the 9309 Model 002, 7014 Model S00, or 7015 Model R00 rack with up to four Magstar 3590 Tape. The 3590 A60 can also be integrated in a 3494 Tape Library Model D14 Drive Unit and attach up to 10 Magstar 3590 Tape Drives with adjacent frame support.

Note: Intermixing of Tape Drive Models on a 3590 Model A60 is not permitted.

8.2 3490E Model F Tape Subsystem

IBM 3490 Magnetic Tape Subsystem Enhanced Capacity Model F has compatibility, high performance, and high availability.

The 3490E FC0 rack with F11 tape drives is illustrated in Figure 68.

Figure 68. 3490E Model F Magnetic Tape Subsystem

8.2.1 Overview

The 3490E Model F is the magnetic tape subsystem designed to provide compatibility for 3480 or 3490 tape formats. The 3490E Model F uses IBM Cartridge System Tape (CST) and IBM Enhanced Capacity Cartridge System Tape (ECCST) with 18/36-track formats. Using a 36-track head while reading or writing 18 tracks bidirectionally, a 3490E Model F doubles the capacity of a base CST and virtually eliminates the rewind time for a full cartridge.

The 3490E Model F supports ECCST and 36-track recording to provide the highest data capacity per cartridge of any 3480/3490 tape format cartridges. With Improved Data Recording Capability (IDRC) compaction, capacity can be up to 2.4 GB — nearly three times higher than the uncompressed capacity of 3480/3490 cartridges. The increased capacity reduces both the number of cartridges needed and overall storage costs.

See 11.2.10, "Compression" on page 220 for more details about data compression techniques.

Using IDRC and a SCSI interface, effective data throughput can be as high as 6.8 MB/sec. The fast-and-wide SCSI-2 differential interface and high-performance control unit buffering are key factors in the improved performance of the new models. The SCSI-2 attachment gives users the potential for higher effective throughput than that currently provided by parallel channel attachment.

The long cartridge life of ECCST and high quality of the 3490E F Models make them a choice to support critical data.

8.2.2 Specifications

The 3490E F line contains the following models:

3490E Model F00 tape drive is the table-top tape drive. The F00 drive must be used through SCSI interface.

3490E Model F01 tape drive is the stand-alone tape drive with a 10-cartridge Automatic Cartridge Facility (ACF). The Model F01 drive must be used through a SCSI interface.

3490E Model F11 tape drive is the rack-mount tape drive for SCSI attachment. F11 is equipped with a 10-cartridge ACF, which holds a cleaning cartridge that provides on-demand cleaning. When F11 drives are mounted on the FC0 rack, you can use them through the FC0 ESCON attachment.

3490E Model F1A is the drive designed for installation in the 3494 Magstar tape library. You can install one or two F1A drives in any new 3494 L10 or D10 frame on independent sliders that enable the drives to be serviced from the rear. This prevents disruption to the 3494 accessor when the drive is being serviced. (Note that the current frame design means that the F1A cannot be installed in existing L10 frames. The autoload feature is not available on the F1A.)

When configured with the appropriate controller feature codes, the F1A can attach to an S/390, SP, AS/400, or RS/6000 system. The F1A can attach to the S/390 through ESCON or OEMI channels. The F1A can attach to the RS/6000 or SP through ESCON or SCSI, and to the AS/400 through SCSI. You can configure the F1A with up to two channel adapters, which can be ESCON, OEMI, or a combination of both. For SCSI attachment, no controller features are required and one SCSI attachment is provided.

3490E Model FC0 rack, FC0 rack controller features, and F11 tape drive enable you to configure a high-performance tape subsystem. The FC0, when ordered with the appropriate controller features, provides attachment to S/390, RS/6000, SP, and AS/400 systems through ESCON, OEMI, or SCSI adapters.

You can install F11 tape drives in the FC0 rack in one-drive increments. F11 tape drives are native SCSI and, if attached to a SCSI host, require no controller features.

FC0 racks support four F11 tape drives, one control unit feature, and up to two channel adapters. Both can be ESCON, OEMI, or one of each. The FC0 rack controller performs ESCON-SCSI protocol translation. The FC0 rack controller feature can support up to eight F11 tape drives. For an eight drive configuration, a second FC0 is used as an expansion frame. The controller feature is not needed for a second FC0.

The FC0 rack, with appropriate controller feature codes, and the F11 tape drive significantly improve tape subsystem throughput. An ESCON-attached FC0 with four F11 tape drives has a maximum uncompacted subsystem throughput of 11 MB/sec compared to 3 MB/sec for an ESCON-attached 3490E A10/B40 configuration. The improved throughput is a result of having the tape drives reside on a shared SCSI bus with a bandwidth of 20 MB/sec.

Hardware specifications

These are the hardware specifications:

- IDRC compression: Standard

- Recording format: 18/36-track format

  Note: On 18-track format, both read and write are available for RS/6000, but only read is available for other platforms.

- 10-cartridge auto loader is standard in F01, F11

- SCSI-II differential interface: Standard

- Maximum compacted data rate: 6.8 MB/sec

- Cartridge capacity: 0.8 GB

- Cartridge capacity (with IDRC): Up to 2.4 GB

- Total capacity (with ACF): Up to 24 GB per F01, F11

- Number of F11 drives: up to four per FC0 rack, up to eight per FC0 controller feature

- Number of ESCON channels: up to four per FC0 controller feature

Minimum software requirements

The software requirements for 3490E F Models are as follows:

For S/390 and S/370 servers:

- MVS/ESA SP Version 4 Release 2.0 with MVS/DFP 3.3.0

- VM/ESA Version 1 Release 1.1

- VSE/ESA Version 1 Release 2.0

For open systems servers:

- RS/6000, RS/6000 SP running AIX/6000 Version 3.2.5 with PTF or Version 4.1.1

- AS/400 running OS/400 Version 3.1

- Sun running Solaris 2.x

8.3 3494 Tape Library Data Server

The IBM Magstar 3494 tape library data server automates the retrieval, storage, and control of 3590/3490 1/2 inch cartridge system tapes, 3490Eenhanced capacity cartridge system tapes, and Magstar 3590 tape cartridges.

The Magstar 3494 tape library is illustrated in Figure 69.
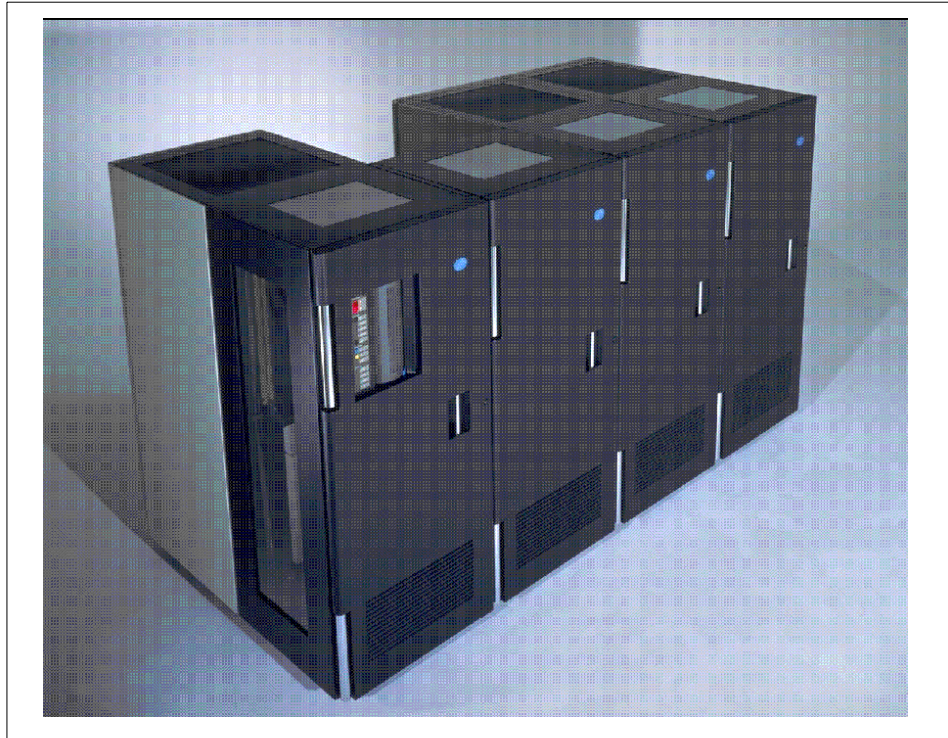
Figure 69. 3494 Tape Library Data Server

8.3.1 Overview

The IBM Magstar 3494 tape library is a modular system that contains a cartridge accessor, a library manager, and space for tape cartridges, with the option for Magstar 3590, 3490E tape drives, and a Magstar Virtual Tape Server.

The Magstar 3494 tape library provides a wide range of configuration flexibility. You can start small and grow in an affordable and incremental manner, beginning with one control unit and adding drive units, storage units, and a Magstar Virtual Tape Server unit, to a maximum configuration of 16 units. The stand-alone Model B18 Virtual Tape Server is not counted in these totals. Configurations of 1 to 8, 10, 12, or 16 frames are allowed, without a Model HA1 High Availability Unit attached to the 3494. When a Model HA1 High Availability Unit is attached, configurations of 3, 4, 6, 8, 10, 12, and 16 frames are allowed (not counting the two service bay frames). Magstar 3590

and 3490E tape drives can be intermixed in the same library along with their tape cartridges.

See 8.1, "3590 High Performance Tape Subsystem" on page 153, and also 8.2, "3490E Model F Tape Subsystem" on page 157 for more detail.

See *IBM Magstar 3494 Tape Libraries: A Practical Guide*, SG24-4632, for more details about the 3494 tape library.

8.3.2 Key features

Magstar 3494 tape library data server includes these key features, described in the following sections:

- 8.3.2.1, "High performance" on page 162
- 8.3.2.2, "High availability" on page 162
- 8.3.2.3, "Wide platform connectivity" on page 163
- 8.3.2.4, "Affinity to storage management software" on page 163

8.3.2.1 High performance

Magstar 3494 is designed for high performance. With a one-unit library configuration, an average mount-access time for the Magstar 3494 cartridge accessor is 7 seconds, with accessor capability up to 265 exchanges per hour with a single gripper, and up to 305 exchanges per hour with a dual gripper. The exchange capability of the accessor typically decreases as the number of frame units increases.

Inventory time is about 4 minutes for each library unit (control unit, drive unit, or storage unit).

Magstar 3590 has a maximum data transfer rate of 14MB/sec in uncompacted mode. The maximum instantaneous data rate is 34MB/sec on an Ultra SCSI interface and 17 MB/sec on an ESCON channel.

See 8.1, "3590 High Performance Tape Subsystem" on page 153 for more detail about the 3590 tape drive.

8.3.2.2 High availability

The Magstar 3494 High Availability Model 1 (HA1) eliminates the accessor and library manager as single points of failure. The HA1 consists of a second accessor, library manager, and all related hardware. The second accessor and library manager operate in hot standby mode. The HA1 is designed for users who require the highest levels of availability from their tape automation subsystem. The HA1 is available with or without the Virtual Tape Server feature.

Magstar 3494's cartridge accessor with the dual gripper feature provides increased performance and availability.

8.3.2.3 Wide platform connectivity
The Magstar 3494 attaches to multiple systems, including AS/400, RS/6000, RS/6000 SP, ES/9000, ES/3090, S/390, and selected UNIX systems from Convex, CRAY Research, Inc., and Sun Microsystems.

8.3.2.4 Affinity to storage management software
In an OS/390 environment, you must use either DFSMS/MVS or Basic Tape Library Support (BTLS) for tape library management. DFSMS/MVS co-operates the Magstar 3494 tape library, which provides efficient tape library management.

See Chapter 13, "Data Facility Storage Management Subsystem" on page 253 for more details.

In UNIX environment, you can use Tivoli Storage Manager which provides the useful open-systems data backup solution.

See Chapter 12, "Tivoli Storage Manager" on page 223 for more details.

8.3.3 Specifications
The Magstar 3494 tape library consists of the following components.

3494 L1x control unit — This unit contains the library manager, console, cartridge accessor, optional convenience I/O stations, optional tape units, barcode reader, optional ESCON controller, and up to 240 cartridge storage cells:

- Model L10 — up to two 3490E tape drives
- Model L12 — up to two Magstar 3590 tape drives
- Model L14 — up to two ESCON attached Magstar 3590 tape drives.

  Note: Model L14 and L12 frames can be attached to a 3590 Model A60 Controller in an adjacent 3494 Model D14 frame when utilizing the Adjacent Frame Support feature.

The Magstar 3590 Model E1A Tape Drive cannot be attached to the same 3590 Model A60 Controller with the Magstar 3590 Model B1A Tape Drive.

3494 D1x drive unit — This unit provides the capability for adding tape units and up to 400 cartridge storage cells:

- Model D10 — up to two 3490E tape drives

- Model D12 — up to six SCSI-attached Magstar 3590 tape drives
- Model D14 — up to four Magstar 3590 tape drives and an ESCON A60 controller.

  Note: The Model D12 frame can be attached to a 3590 Model A60 Controller in an adjacent 3494 Model D14 frame when utilizing Adjacent Frame Support.

Magstar 3590 Model E1A Tape Drive cannot be attached to the same 3590 Model A60 Controller with Magstar 3590 Model B1A Tape Drive.

3494 S10 storage unit — This provides the capability for adding cartridge storage:

- Model S10 — up to 400 cartridge cells

3494 B1x Virtual Tape Server control unit — This unit is designed to stack multiple host-created volumes onto a Magstar 3590 cartridge. Model B1x contains Magstar Virtual Tape Server controller, fault-tolerant RAID disk cache storage, and up to 400 cartridge storage cells. Model B18 requires an Model D12 that has cartridge storage, and contains from three to six Magstar 3590 drives that are used by the Virtual Tape Server.

B1x and D12 can be in any position of a Magstar 3494 system configuration unit that can coexist with drive frame units containing 3490E or Magstar 3590 tape drives. Magstar Virtual Tape Server connects to OS/390 systems through one or two ESCON attachments.

See 8.4, "Virtual Tape Server" on page 166 for more details about the Virtual Tape Server.

Hardware specifications

The hardware specifications are as follows:

- Number of tape cartridges: up to 6240 per 3494 library
- Total capacity: up to 748 TB per 3494 library
- Average mount-access time: 7 seconds
- Number of frames: up to 16 (without HA1 frames)
- Number of ESCON channels: up to 4

See 8.1, "3590 High Performance Tape Subsystem" on page 153; and 8.2, "3490E Model F Tape Subsystem" on page 157, for the specifications of tape drives for 3494 tape library.

The requirements for the Magstar 3494 tape library are described in the following section.

Attachment requirements

Host operating systems send data to Magstar 3590 tape drives through SCSI or ESCON attachment with the Magstar 3590E having the addition of fibre. The 3490E tape drives through ESCON, parallel, or SCSI attachment. The host system software communicates library commands to the Magstar 3494 Library Manager, such as volume/drive mount commands. This communication is done through LAN or RS-232 attachment, or through the ESCON or parallel channel controller.

- **ES/9000** — The Magstar 3494 supports all models of ES/9000 9221, 9121, and 9021 systems and all ESA-capable S/370 and S/390 systems. Programming support is provided for MVS/ESA (using either DFSMS/MVS or Basic Tape Library Support), TPF, VSE/ESA, and VM/ESA operating environments. VM/ESA provides mount and demount tape support for guest operating systems such as VSE/ESA and MVS/ESA.

  DFSMS/MVS, system managed tape, and the Magstar 3494 together provide full media independence in the library.

- **AS/400** — AS/400 support for the 3494 is provided through OS/400 and Backup Recovery and Media Services/400 (BRM Services/400).

- **RS/6000** — Support is provided for AIX/6000 operating environments through a programming interface for applications and user commands.

- **Convex, CRAY, and Sun Microsystems** — Convex supports the Magstar 3494 with the Magstar 3590 B1A only and provides all the necessary device drivers for the connection.

  Magstar 3494 supports the Magstar 3590 and 3490E tape drives in the CRAY environment and requires Feature 5212 and an RS/6000 operating as a mount server.

  Magstar 3494 supports Magstar 3590 and 3490E tape drives in the Sun environment and requires Feature 9204.

  Silicon Graphics intends to release a version of its IRIX operating system that supports the Magstar 3590 Tape Subsystem and the Magstar 3494 Tape Library.

See *IBM Magstar 3494 Tape Libraries: A Practical Guide*, SG24-4632, for more details about open systems attachments.

8.4 Virtual Tape Server

The Virtual Tape Server (VTS) provides efficient use of tape library resources, which reduces storage costs. See Figure 70.

Figure 70. The Virtual Tape Server

8.4.1 Overview

By stacking multiple data volumes onto the high-capacity 3590 cartridge, the VTS (shown in Figure 70) leverages the capacity of Magstar 3590 tape cartridges, lowering tape operation costs by reducing floor space, as well as reducing the number of physical tape drives and cartridges. VTS also reduces tape operation costs through transparent (from host) and effective use of tape resources, reducing the number of cartridges, tape drives, and automation systems required to support a workload.

VTS is an IBM Seascape solution, designed to stack multiple host-created 3490E tape volumes on a Magstar 3590 cartridge. The 3494 VTS provides attachment to both S/390 and selected Open Systems environments.

- **IBM Magstar 3590 tape drives** (See 8.1, "3590 High Performance Tape Subsystem" on page 153 for more details.)

- **IBM RAID serial disk storage** (See 6.2.2, "Serial Storage Architecture" on page 83 for more details.)

- **IBM Magstar 3494 tape library** (See 8.3, "3494 Tape Library Data Server" on page 160 for more details.)

- **IBM RS/6000**

- **Robust, proven storage management software** (See Chapter 12, "Tivoli Storage Manager" on page 223 for more detail.)

VTS is a multiplatform consolidation product. To find out more about IBM enterprise storage consolidation, see 16.1, "Understanding enterprise storage administration" on page 305.

The Basic hardware diagram of VTS is illustrated in Figure 71.
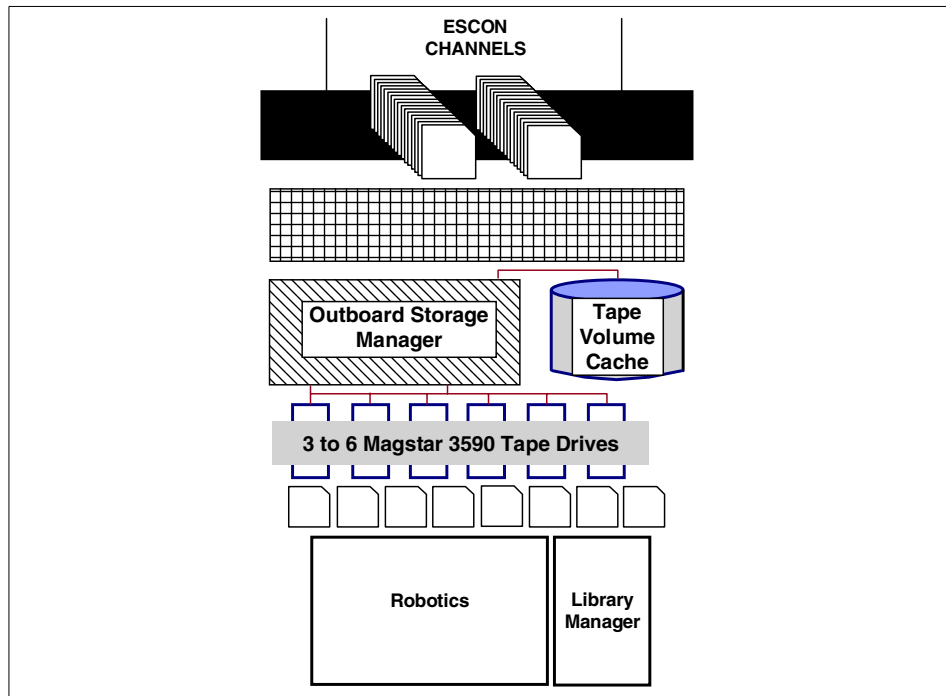
Figure 71. VTS hardware structure

VTS introduces the following concepts:

- **Tape volume cache** — Fault-tolerant RAID disks are managed by VTS intelligent storage management software as tape volume cache. Access to all data is through the tape volume cache, which extends many of the performance benefits of cache disk to tape. For example, creating new tape volumes (nonspecific mounts) is done solely to the tape volume cache.

  The cache can hold hundreds of virtual volumes. The content of the cache is managed to retain the most recently accessed virtual volumes so that numerous subsequent mount requests can be satisfied very quickly from the cache, similar to a cache request on DASD. If a requested volume is not present in the cache, the required Magstar 3590 cartridge is mounted, and the logical volume is moved back into the cache from a stacked volume.

  With the 3494 B18 VTS control unit, tape volume cache is up to 5184 GB when the Performance Accelerator feature is installed.

- **Virtual tape drive** — As viewed by host software, all the functions of a 3490E drive are provided. Data transferred between hosts and virtual devices is stored in the tape volume cache.

  With the B18, virtual tape drive capacity is up to 64 drives per VTS and 128 virtual drives with two Magstar Virtual Tape Servers in a single library configuration.

- **Virtual volume** —As viewed by host software, this represents a 3490E cartridge where host-accessible data volumes are mounted on virtual devices. Virtual volumes are stored in the tape volume cache and subsequently moved to stacked volumes.

- **Logical volume** — Data volumes stored on a stacked volume are logical volumes. For host access, a logical volume must be moved to the tape volume cache.

  With the B18, logical volume is up to 250,000 volumes per 3494 tape library.

- **Stacked volume** — Stacked volumes are Magstar 3590 volumes that contain several logical volumes. After a virtual volume is unloaded, it is copied onto a stacked volume. The virtual volume then remains in the cache until its space is required for another virtual volume. The content of stacked volumes is managed by VTS such that partially full stacked volumes are consolidated to free up space.

- **3490E emulation** — From a host perspective, data is processed as if it resides on actual 3490E devices and cartridges. This representation of tape devices enables transparent use of Magstar 3590 tape technology. It also enables use of VTS without requiring installation of new software releases. Data is stored on disk as images of either virtual Cartridge System Tape or Enhanced Capacity Cartridge System Tape. All 3490E-type commands are supported. Tape motion commands are translated into disk commands, resulting in response times much faster than in conventional tape drives.

See *Magstar Virtual Tape Server Implementation Guide* (SG24-2229) for more detail.

8.4.2 Specifications

VTS consists of the following components:

- **3494 B1x VTS control unit** contains a tape volume cache, VTS controller, an RS/6000,and power control and distribution. It can coexist with 3490E and native 3590 subsystems in a 3494 library. One 3494 tape library can contain two VTSs, B16/B18 or 2 B18s. The Model B18 must reside within 14 m from the 3494 Model D12. Model B16 supports up to two ESCON channel attachments. Model B18 supports up to four ESCON channel attachments.

  - Model B18 — The B18 is packaged as a stand alone unit that can be a maximum of 14 meters from the associated 3494 drive unit frame. The B18 provides a maximum of 64 virtual devices and up to 250,000 logical volumes per VTS.

    Note: An existing 3494 Model B16 Virtual Tape Server can be converted to a Model B18 Virtual Tape Server with the EHPO feature.

- **3494 D12 drive unit** contains three to six 3590-B1A tape drives used by VTS control unit and cartridge storage cells.

Optional 3494 VTS components

- Model AX0 — The 3494 Model AX0 Virtual Tape Controller interconnects with two 3494 Model B18 VTSs having the Peer-to-Peer Copy features in a Peer-to-Peer Virtual Tape Server configuration. Each Model AX0 provides two ESCON host attachments for the Peer-to-Peer Virtual Tape Server.

- Model CX0 — The IBM 3494 Model CX0 Auxiliary Frame provides the housing and power for two or four 3494 Model AX0 Virtual Tape Controllers. There are two power control compartments, each with its own power cord to allow connection to two power sources.

Note: The Magstar 3494 Model CXO and AX0 Virtual Tape Controllers are part of a Peer-to-Peer feature of the VTS.

8.4.3 VTS performance

- A performance accelerator

An enhanced controller containing a four-way SMP processor with an enhanced file management system that improves sustained and high performance. Also, it reduces answer time for virtual volume recalls, so that there is more time for host data transfer.

- ESCON Channels

Eight independent ESCON channel adapters, which optimize efficiency. The channel compression option offers more disk capacity and better performance.

- Magstar 3590E

Fast drives and 20 GB uncompressed cartridges, which results in fewer cartridges needed. The use of 3590 Extended Performance Cartridges are planned in the future. (See 8.1, "3590 High Performance Tape Subsystem" on page 153.)

8.4.4 VTS features

- Peer-to-Peer VTS

The Magstar Peer-to-Peer Virtual Tape Server automatically creates a copy of any newly created or updated tape volume in both Virtual Tape Servers. This process is performed transparently to the customer application and with no host processor resources required. Either volume copy can then be used to satisfy a specific customer mount. This copy of all new or updated virtual volumes can be created using one of two possible modes of operation:

    - **Immediate Copy** — creates a copy of the logical volume in the companion connected Virtual Tape Server prior to completion of a Rewind/Unload command. This mode provides the highest level of data protection.

    - **Deferred Copy** — creates a copy of the logical volume in the companion connected Virtual Tape Server as activity permits after receiving a Rewind/Unload command. This mode provides protection that is superior to most currently available backup schemes. (See "Magstar Peer-to-Peer Virtual Tape Server Planning and Implementation Guide, SG24-6115-00.)

- Import/Export

  The 3494 B18 VTS data portability is provided by a set of functions called Import and Export. The Import and Export functions allow you to move data between VTSs without having to copy the logical volumes back through the host channels, or lose the VTS benefits of volume stacking, which exploits the storage capacity of the 3590 cartridge. Exported logical volumes are not accessible any longer in the VTS after the Export, and become accessible in a VTS after an Import.

  Note: The Import/Export function is not supported with Peer-to-Peer VTS and is disabled during Peer-to-Peer VTS migration.

- "Call Home" Service Support

  The "Call Home" function, when activated, automatically opens a service alert when a problem occurs with the 3494 Model B18 or AX0 or a Magstar 3590 Model A60 or A50 Controller in a 3494 Tape Library. The "Call Home" process can also be initiated by a user from the 3494 Library Manager operator panel for a 3494 Model B18 or AX0 or a Magstar 3590 Model A60 or A50 Controller installed in a 3494 Tape Library.

- SAN connection

  The IBM 2108 Storage Area Network Data Gateway Model G07 can be used to connect Fibre Channel capable supported servers to the 3494 Model B18 VTS configured with one or more SCSI Host Attachments.

- IBM StorWatch Specialist for the 3494 Tape Library

  Two Specialists are available on the 3494:

  The 3494 Tape Library Specialist is a system- independent, Web-based user interface to the 3494 library manager, which enables you to monitor the library manager and attached Virtual Tape Servers from a remote location. Connection to customer LANs will be through the Token-Ring Adapter or the Ethernet Adapter on the L1x Control Unit frame.

  The 3494 Peer-to-Peer VTS Specialist is the Web user interface used to view the current status and configuration of the Peer-to-Peer VTS. This Specialist requires an ethernet network which is connected to the 3494 Model AX0 Virtual Tape Controllers.

Hardware specifications

The hardware specifications are as follows:

- Number of 3590 tape drives: three to six per VTS
- Number of virtual tape drives: up to 64 per B18,or 128 with 2 VTSs
- Number of logical volumes: up to 250,000 per 3494 tape library
- Number of ESCON channels: up to 8 per B18

Minimum software requirements

The minimum software requirements are as follows:

- MVS/ESA SP 4.3
    - DFSMS/MVS 1.1 or later for system-managed tape support
    - MVS/DFP 3.3 BTLS for basic tape library support
- VM/ESA Version 1.2 or later
- VSE/ESA Version 1.3 or later when running as a guest of VM/ESA Version 1.2 or later.

8.4.5 Future enhancements

- Up to 12 3590 drives per VTS
- StorWatch Specialist enhancements
- Support for blocksizes larger than 64 KB with OS/390 BSAM and QSAM access methods
- Management of the retention of expired virtual volumes
- Use of 3590 Extended Performance Cartridges

8.5 3570 Magstar MP Tape Subsystem

The IBM 3570 Magstar Multi Purpose Tape Subsystem is an integrated tape library, enhanced to deliver higher levels of performance to SCSI attached systems. Larger amounts of data can be backed up in less time, significantly improving backup window performance. See Figure 72.

Figure 72. The 3570 Magstar Multi Purpose Tape Subsystem

8.5.1 Overview

The 3570 Magstar MP (shown in Figure 72) is based on the IBM 3590 Magstar technology. It is designed to provide a reliable automated tape solution to the major midrange SCSI-attached systems, such as AS/400, RS/6000, HP-UX, Sun Solaris, and Microsoft Windows NT.

It handles a new cartridge with a different shape, uses a different format from previous IBM standards, and is not compatible with them.

The 3570 Magstar MP tape library models use a cassette loading and transport mechanism (priority slot) to automatically transport the tape cassettes to and from the cassette magazines and the tape drive.

8.5.2 Key features

The 3570 Magstar MP key features include:

• High capacity and performance for midrange systems

  Data retrieval actions taking minutes on other tape media take only seconds on 3570 Magstar MP. At last, tape can be legitimately called *near on-line storage*. The 3570 Magstar MP drive, library, and cartridge allow:

- 7 MB/sec native sustained drive data rates; 50 GB/hr data rates per drive with maximum compression

- Midpoint cartridge load, and average access of 4 seconds to first byte

- Cartridge capacity of 21 GB using 3:1 compression when you use a new C-XL format cartridge

- Fast average cartridge exchange time of 6 seconds, and a barcode reader for a fast identification of the cartridge in the library

- The library operates with two 10-cartridge magazines, providing from 100 to 300 GB of data accessible to the host system.

• High reliability using Magstar technology

The tape medium itself is a strong metal case offering a dust-free environment that has very reliable data retention capabilities. At no time is the tape removed from the cartridge, so the possibility of attrition of the medium through contact with wheels, pinions, channels, or recording drums is minimized. To ensure the accurate and consistent reading of the 128 tracks of data, the latest generation of IBM's magneto resistive (MR) heads is employed. A cleaner cartridge is stored in the library for automatic cleaning of the tape drive.

• Designed for automation

The 3570 Magstar MP is designed for automation while fulfilling all the requirements of a basic backup device. It offers a sequential mode for traditional tape processing applications and a random mode for applications requiring direct access to data.

8.5.3 Specifications

There are five IBM Magstar MP models intended to be housed in a standard rack or for desk-top use. All five models use one SCSI cable:

• 3570-C00 — a single-drive desk-top unit

• 3570-C01 — a 20 cartridge desk-top library, containing one drive

• 3570-C02 — a 20 cartridge desk-top library, containing two drives

• 3570-C11 — a 20 cartridge rack mountable library, containing one drive

• 3570-C12 — a 20 cartridge rack mountable library, containing two drives

• 3570-C21 — a 20 cartridge Netfinity rack mountable library, containing one drive

• 3570-C22 — a 20 cartridge Netfinity rack mountable library, containing two drives

An IBM Magstar MP 3570-B is available in these five models as well. The IBM Magstar MP 3570-B provides the same reliability and capacity as the 3570-C with a 2.2 MB/sec native sustained drive data rate.

The library units may be set to automatic, manual, or random modes from the operator panel:

- Automatic mode — Automatic (sequential) mode allows the operator to keep the drive continually fed with cartridges from the magazines. After being enabled, the transport mechanism automatically loads and unloads cartridges until all of the cartridges in the magazine are processed.

- Manual mode — Manual mode allows the same functions to occur as in the automatic mode, except that manual mode allows the operator to initiate each load operation, by pressing Start from the *Set Library Mode* menu. Only one cartridge is processed for each operator action.

- Random mode — Random mode allows the IBM 3570 Magstar MP to act as a self-contained library of up to 19 cartridges, controlled entirely by SCSI initiators. In addition, access is provided to Cartridge 20 through the priority cell.

  A fully intelligent tape library with automated data migration to and from the library can be achieved by the addition of a storage management software solution such as Tivoli Storage Manager or Backup Recovery and Media Services/400 (BRMS/400) for the AS/400. For standard operation with Magstar MP under a Windows NT environment, a library software manager like Tivoli Storage Manager is necessary.

  For more information, see *The IBM Magstar Virtual Tape Server and Enhancements to Magstar: New Era in Tape*, SG24-4917.

8.6 Linear Tape-Open (LTO)

The Linear Tape-Open (LTO) program is a joint initiative of IBM, Hewlett-Packard, and Seagate Technology. In 1997, the three companies set out to enable the development of best-of-breed tape storage products by consolidating state-of-the-art technologies from numerous sources, and in November of that year produced a joint press release about LTO. The three technology provider companies have all made significant contributions of time and expertise to the definition of the LTO format specifications. (See *The IBM LTO Ultrium Tape Libraries Guide*, SG24-5946.)

An implementation of LTO technology, the Ultrium tape format, is shown in Figure 73.

Figure 73. The IBM Ultrium family of tapes and libraries

8.6.1 Overview

The Ultrium tape format is the implementation of LTO technology optimized for high capacity and performance with outstanding reliability, in either a stand-alone or an automated environment. The Ultrium tape format uses a single reel cartridge to maximize capacity. It is ideally suited for backup, restore, and archive applications. The Ultrium drive reads and writes data using 384 data tracks per cartridge. Servo tracks (also called bands) enable accurate positioning of the tape drive head over the data track. The servo bands are written at time of manufacture. The IBM LTO Ultrium common sub-assembly Generation 1 (sometimes and unofficially referred to as "the brick") is a high performance, high capacity tape drive. This basic unit is the first generation in the IBM LTO series of products and is used throughout the entire IBM Ultrium family of tapes and libraries.

8.6.2 The IBM LTO Ultrium family of tapes and libraries

The IBM Ultrium family of tapes and libraries (pictured in Figure 73 on page 176) comprises four different product offerings, ranging from a stand-alone unit to a highly-scalable automated library. They are all based around a common tape drive sub-assembly packaged in different robotic and stand-alone environments.

- The **IBM 3580** is a stand-alone desk-top single drive unit without autoloader. Tapes are mounted manually one at a time.

- The **IBM 3581** is also a single drive unit but has a 7 cartridge autoloader within the device. It is a stand-alone desk-top unit; however, optional additional hardware is available for installation in a standard 19-inch rack.

- The **IBM 3583** is a small robotic library, accommodating from a minimum of 1 up to a maximum of 6 drives and providing space for up to 72 Ultrium cartridges. The IBM 3583 may be stand-alone or, with an optional feature, housed in a rack.

- The **IBM 3584** is a larger modular library with the potential to house a maximum of 72 tape drives in the maximum 6 frames. There is a trade-off between cartridge capacity and installed drives, so that a fully-configured library with 72 drives and a 10-cartridge I/O station has a cartridge capacity of 2207; with a minimal drive configuration, the cartridge capacity can go up to a maximum of 2481.

Storage and tape management for the IBM LTO Ultrium family is provided by software such as Tivoli Storage Manager and other similar software offerings.

These products, while all part of the same family, are distinct offerings. The drive and cartridge technology and formats are the same, and the cartridges are interchangeable between the libraries as dictated by the LTO format standards. However, the machine types (for example, 3583 and 3584) are not upgradeable from one to another, nor can the Ultrium drives be exchanged between different libraries. Thus, if you purchase an IBM 3583 library with six drives, and later on you install an IBM 3584, you cannot transfer the Ultrium drives from the 3583 to the 3584.

8.6.3 Key features

- Tape capacity up to 100 GB native per cartridge (200 GB with 2:1 compression, 3580)

- Native data capacities up to 248.1 TB (up to 496.2 TB using 2:1 compression, 3584)

- Data transfer rate up to 15 MB per second (uncompressed), up to 30 MB per second with 2:1 compression

- Low-Voltage Differential (LVD) Ultra2 SCSI attachment that connects to LVD fast/wide adapters

- High-Voltage Differential (HVD) Ultra SCSI attachment that connects to HVD fast/wide adapters

- Adheres to Linear Tape-Open (LTO) specifications

8.6.4 Specifications

- 4 models, single drive stand alone unit to 72 drive automated library.

- The SCSI interface (HVD or LVD) or Fibre attachment via Fibre Channel server adapters through selected IBM Storage Area Network Data Gateway products. Future plans include Fibre Channel-Arbitrated Loop (FC-AL) attachment.

- IBM Ultrium drives share best-of-breed technology with advanced multi-track recording capabilities, magneto-resistive (MR) head and servo technologies, and high recording densities. Error correction code (ECC) function helps provide data integrity.

8.7 3575 Magstar MP Tape Library Data Server

The IBM 3575 Magstar Multi Purpose (MP) Tape Library Data Server (see Figure 74) has been enhanced with the addition of a patented new multipath architecture, supporting sharing by homogeneous and heterogeneous hosts.

Figure 74. 3575 Magstar MP Tape Library Data Server models

8.7.1 Overview

The 3575 Magstar MP Tape Library Data Server provides a midrange and network-server tape storage solution combining reliable, field-proven, automated tape handling and storage.

The 3575 Magstar MP Tape Library Data Server uses the high-performance IBM 3570 MP Model C tape drives, discussed in 8.5, "3570 Magstar MP Tape Subsystem" on page 172, and therefore can be connected to the same hosts.

You can tailor the library to match your system capacity and performance needs, from 300 GB to 1.6 TB (900 GB to 4.8 TB with 3:1 compression), using up to six Magstar MP tape drives.

The 3575 can be connected to up to six SCSI-attached hosts, in up to three logical libraries, sharing a common tape cartridge inventory and robotics.

8.7.2 Key features

The IBM Magstar MP 3575 Tape Library provides Magstar MP tape reliability. Other features include:

- Support for up to three logical libraries

- A multipath architecture, with a multihost attachment for up to six SCSI control paths
- Library cartridge move time of less than 4 seconds
- Improved bulk loading enables 14 storage slots to be allocated as user-selectable bulk I/O slots.

For open system architectures, this means that this library can be connected up to six different servers with different operative systems with SCSI attachment. All your servers AS400, UNIX systems, Windows NT, and other platforms, can be connected to the IBM Magstar MP 3575, thus reducing the number of tape units spread around your network.

Figure 75 represents the different IBM Magstar MP 3575 models. The arrows indicate model upgrades. L12 can be upgraded to L18; likewise, L24 can be upgraded to L32.

Magstar MP - 3575 Models

	L06	L12	L18	L24	L32
# Drives	1 or 2	1 - 4	1 - 6	1 - 6	1 - 6
#Slots	60	120	180	240	324
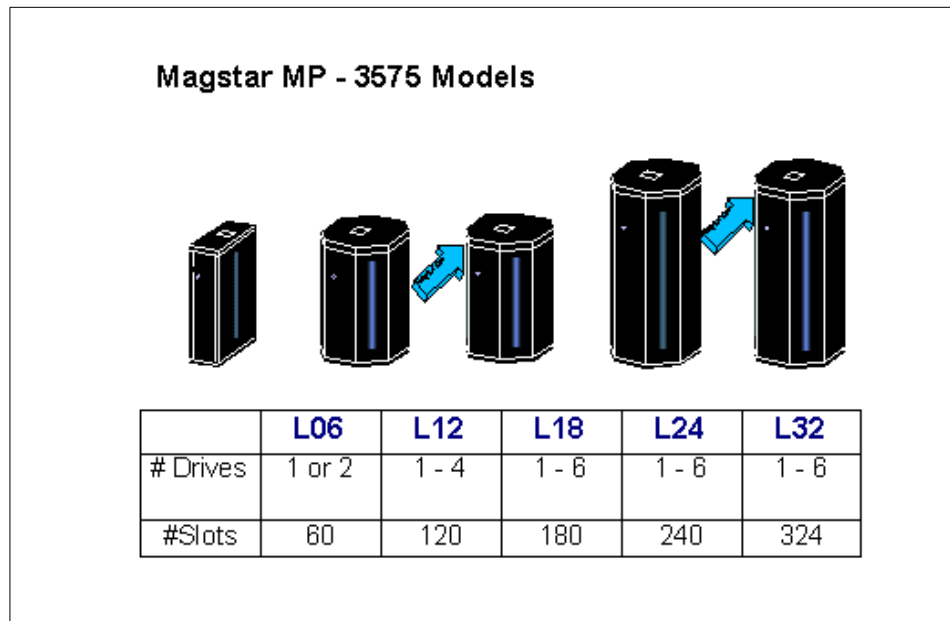
Figure 75. Magstar 3575 models

8.7.3 Specifications

The Magstar MP 3575 tape library models are compact, high-capacity, integrated storage devices that are available as stand-alone units. Five models are available; see Table 6.

Table 6. 3575 Magstar MP Tape Library Design

Model	L06	L12	L18	L24	L32
Native capacity (MB)	300	600	900	1200	1600
Drives	1-2	1-4	1-6	1-6	1-6
Number of cartridges	60	120	180	240	324
Number of logical libraries	1	1-2	1-3	1-3	1-3

The multipath architecture permits the logical partitioning of one physical library into several logical libraries. Each logical library consists of tape drives, cartridge slots, and cartridges that are unique to each logical library. The I/O station and the transport mechanism are shared between the logical libraries on a first-come, first-served basis. With the 3575, each drive and each cartridge belongs to one library. In case of using more attached hosts for one library, the hosts must provide arbitration for the cartridge inventory.

The 3575 also provides a drive auto-cleaning function with automatic detection, and a bar-code reader which improves inventory times.

The following host software applications support the 3575 Magstar MP tape library: IBM ADSM (AIX, Windows NT, AS/400), IBM BRMS/400, IBM NetTAPE for AIX, CA/Cheyenne ARCserve, SCH REELlibrarian, SCH REELbackup, SCH Robot and HYPERtape, Legato Systems NetWorker, Spectra Logic Alexandria, VERITAS Media Librarian (VML) and NetBackup.

For more information, see: *Magstar MP 3575 Tape Library Data Server: Multiplatform Implementation*, SG24-4983.

8.8 3995 Optical Library Data Server

The IBM 3995 Optical Library Data Server family (see Figure 76) introduces a new level in the storage hierarchy by offering a unique balance of capacity, performance, and cost.

Figure 76. IBM 3995 Optical Library C-Series

8.8.1 Overview

The IBM 3995 is a library, based upon industry-standard optical technology.

By providing economical, direct access storage, the 3995 is ideally suited for infrequently accessed information that currently resides on magnetic disk, tape, paper, microfiche, or microfilm. Ideal applications include: document imaging, Computer Output to Laser Disk (COLD), report management, hierarchical storage management, and backup/archive.

The 3995 supports SCSI-attachments for the Open Systems environment, AS/400-attachment, S/390-attachment, and LAN-attachment.

8.8.2 Key features

The IBM 3995 key features include:

- Long life for online and for archival data, and the option for permanent recording
- Cross-platform solutions: Open Systems, S/390, AS/400 and LAN environments

8.8.3 Specifications

The 3995 is based upon 5.25-inch, industry-standard Extended Multifunction optical drives which enable support for magneto-optical (MO) rewritable, Permanent Write Once/Read Many (WORM), and Continuous Composite WORM (CCW) recording technologies. Each cartridge has a capacity of 2.6 GB.

The 3995 offers the models shown in Table 7, depending on the operating environments.

Table 7. 3995 C-series optical libraries

MODEL	CX0	CX2	CX4	CX6	CX8
Cartridges	20	52	104	156	258
Drives	1 or 2	2	2 or 4	4 or 6	4 or 6
Capacity	104GB	270GB	540GB	811GB	1.34TB
Attachments: **X = 6** **X = 2** **X = 4** **X = 3** **X =1 expansion**	SCSI LAN AS/400 -- --	SCSI LAN AS/400 S/390 S/390	SCSI LAN AS/400 S/390 --	SCSI LAN AS/400 S/390 S/390	SCSI LAN AS/400 S/390 S/390
Model Upgrades	-	-	From	To	-

Chapter 9. SAN storage solutions

This chapter addresses other storage solutions that should be considered for your storage configuration.

9.1 SAN Data Gateway for serial disk

This section provides information related to the SAN Data Gateway solution available for serial disk.

9.1.1 Overview

SAN Data Gateway for serial disk IBM 2108 Model 020 (see Figure 77) provides new or existing UNIX and Windows NT host systems attachment to 7133 Serial Disk systems and serial storage technology, where native connection is not available.

Figure 77. SAN Data Gateway Model 020
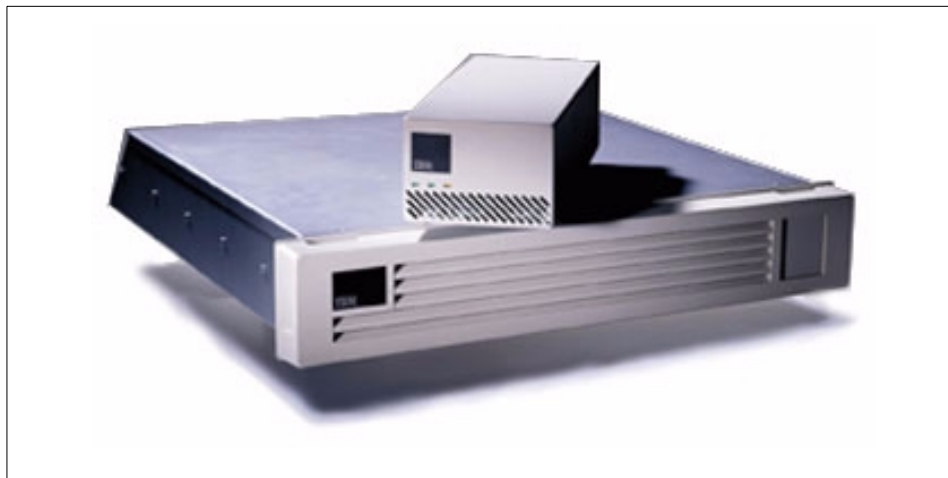
9.1.2 Key features

The SAN Data Gateway Model 020:

- Enables you to attach Ultra SCSI and SCSI-based host servers to the IBM 7133 Serial Disk System, the 7131 Model 405 Multi-Storage Tower, and the 3527 SSA Storage Subsystem.

- Provides multi-host, highly available data-sharing solutions — more than 2 terabytes of serial disk storage can be attached on a single Gateway

S20 loop, and up to eight Gateway Model S20s can be configured on a loop.

- Offers increased data availability, performance, and protection — the SAN Data Gateway Model S20 may replicate data across serial disk systems and simultaneously mirror two or three copies of data, independent of the host system.

- Includes the StorWatch SAN Data Gateway S20 Specialist, an NT-based management tool that provides configuration and service functions. The Specialist enables you to define RAID1 groups and disk concatenation, and supports splitting off a copy of a mirrored group for backup or processing on another system.

9.1.3 Specifications

The SAN Data Gateway Model S20 is a compact, self-contained unit. It provides one wide UltraSCSI Differential port and two SSA ports. The UltraSCSI Differential port has two 68-pin HD female connectors. An external terminator is provided on one of the SCSI connectors. Both fast/wide SCSI-2 Differential and fast/wide UltraSCSI Differential are supported.

The Gateway will negotiate to the connected host adapter interface, complying with ANSI SCSI-2 specification X3.131-1994. The two SSA ports each have a nine-pin miniD female connector for attaching to a single SSA loop containing 7133 Serial Disk Systems, 7131 Model 405 Multi-Storage Towers, and 3527 SSA Storage Systems.

The SAN Data Gateway Model S20 can be used as a stand-alone tabletop unit, or mounted in a Seascape Solution Rack (2101-100) or equivalent (7014-S00, 7015-R00, 7202-900).

For rack mounting, the optional feature 2000 must be ordered. This feature provides a 2U rack-mountable enclosures which can hold up to four Gateway Model S20s. The Gateway contains a self-regulating power supply that connects to an electrical source, 100V to 240V at 50 to 60HZ.

The SAN Data Gateway requires a SCSI cable for attaching to a host server, and two SSA cables for attaching to the Serial Disk Systems. These cables are customer supplied, and typically ordered with the attaching host server and Serial Disk systems. The total SCSI bus length cannot exceed the differential SCSI limit of 25 meters. The host SCSI bus must be connected only between the host bus adapter and the SAN Data Gateway S20.

Included with the SAN Data Gateway is a CD-ROM that provides the StorWatch SAN Data Gateway S20 Specialist. This StorWatch Specialist runs

on Windows NT systems attached either to the same SSA loop as the SAN Data Gateway S20, or by way of a TCP/IP network connection to the host system attached to the SAN Data Gateway S20.

9.2 Vicom Fibre Channel SLIC Router

The Vicom Fibre Channel SLIC (Serial Loop IntraConnect) Router Model FC-SL (7139-111)enables all IBM 7133, 7131, and 3527 Serial Disk Systems to attach to host systems using Fibre Channel host adapters and drivers.

The Vicom Fibre Channel SLIC Router is shown in Figure 78.
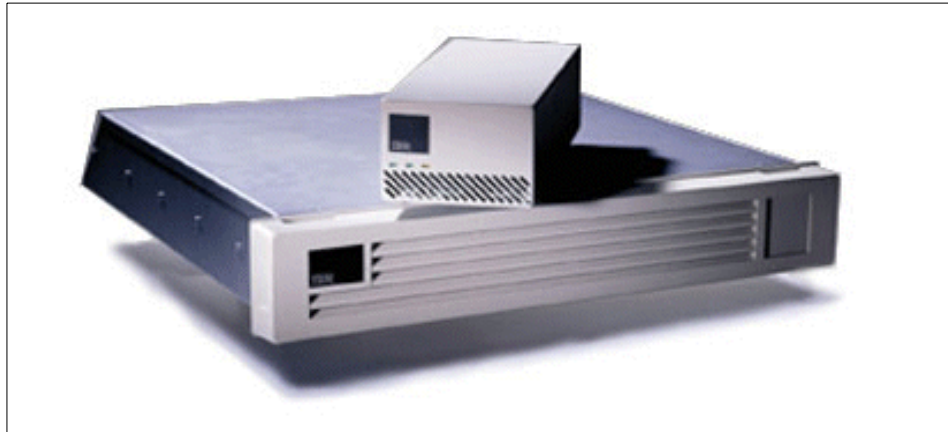
Figure 78. Vicom Fibre Channel SLIC Router

9.2.1 Key features

The Vicom Systems SLIC Router FC-SL provides the following:

- Connectivity for Fibre Channel based servers to IBM's Serial Storage Architecture (SSA)-based disk subsystems

- All the advantages of SSA architecture, including high performance, high reliability, and scalability

- Vicom's Serial Loop IntraConnect (SLIC) technology, which is used worldwide

- Easy-to-use, graphical setup and ongoing management tools

- Advanced functions such as RAID-1 data replication for added data protection and Instant Copy for SSA loop availability

- Support for large amounts of SSA-based disk storage when users take advantage of multiple SLIC routers

9.2.2 Specifications

The IBM 7139 Model 111 has a single Fibre Channel port for connection to an SSA loop consisting of IBM 7131, 7133, and/or 3527 subsystems, easy-to-use tools for initial setup and ongoing management, advanced functions such as RAID-1 data protection and Instant Copy for improved SSA loop availability, and ability to support large amounts of SSA-based disk storage using multiple SLIC routers.

The Fibre Channel SLIC Router replicates data across or within serial disk systems-simultaneously mirroring two or three copies of data without host involvement. With global hot disk sparing, data is automatically rebuilt if a mirrored disk fails. In this way, the Fibre Channel SLIC Router improves performance and data availability while simplifying storage operations.

The Instant Copy function can create a separately addressable copy of mirrored data that can be used for tape backup. After the backup has completed, data is resynchronized with the primary copy.

Configuring and managing is made easy by SLIC Manager.The SLIC Manager is a Windows NT-based management tool that provides configuration and service functions-including mirror group definition, the ability to create composite drives, and Instant Copy disk management. The SLIC Manager can manage multiple Fibre Channel SLIC Routers across the enterprise.

The 7139-111 can be rack mounted in FC 2000 enclosure.This feature provides a mechanical enclosure 2U high - height 3.5 inch (88 mm), width 19.0 inch (483 mm), depth 19.95 inch (507 mm), and weight 22 lbs. (10 kg) which can be mounted on a standard 19 inch rack. It can house up to four 7139-111s. The enclosure is designed to provide a higher level of integration with the current rack-mounted 7133 Serial Disk System. It can be installed in the IBM Seascape Solution Racks (2101-100) or equivalent racks such as 7014-S00,7015-R00 or 7202-900.

For more information, see the Vicom Web page at:

`http://www.vicom.com`

9.3 SAN Data Gateway for UltraSCSI disk and tape storage

This section provides information related to the SAN Data Gateway solution available for UltraSCSI disk and tape storage.

9.3.1 Overview

The Storage Area Network (SAN) Data Gateway connects SCSI and Ultra SCSI storage servers and storage devices to Fibre Channel environments. It attaches new or existing SCSI storage products to storage area network or directly to host systems using an industry-standard Fibre Channel Arbitrated Loop (FC-AL) interface.

9.3.2 Key features

The SAN Data Gateway is an easy-to-manage protocol convertor that configures with one or two FC-AL ports, and provides four Ultra SCSI Differential ports for device attachment. See Figure 12 on page 27.

Initially, it supports attachment of the IBM Enterprise Storage Server (ESS) to Fibre Channel capable Intel-based servers running Microsoft Windows NT, and Fibre Channel capable UNIX servers from Sun. It also supports the attachment of the Magstar MP 3570 Tape Subsystem and the Magstar MP 3575 Tape Library Dataserver to Fibre Channel capable Intel-based servers running Microsoft Windows NT (see Figure 79).

IBM Storage Area Network Gateway integrates Fibre Channel and SCSI technologies

Figure 79. SAN Data Gateway Fibre Channel

9.3.3 Specifications

The SAN Data Gateway is a compact, self-contained unit. There are four standard Ultra SCSI Differential ports, each with a single 68-pin HD connector for attachment of storage servers and devices. The SAN Data Gateway can be configured with one or two FC-AL ports by specifying either feature number 2204 for a one port configuration, or feature number 2205 for a two port configuration.

A SAN Data Gateway that is initially configured with only one FC-AL port (# 2204) can be field upgraded to a two port configuration by ordering feature number 2303. Each FC-AL port provides a duplex shortwave optical interface, and supports attachment of shortwave fiber optic cables with an industry-standard duplex SC connector to specified host adapters or to other storage area network components such as Fibre Channel hubs.

The SAN Data Gateway requires SCSI cables to attach to storage servers and storage devices. Each of the four Ultra SCSI Differential ports can be attached to a SCSI bus up to 25 m in distance, and provides an interface bandwidth of up to 40 MB.

The SCSI connector is a 68-pin HD P-Shell (female) connector. These cables are customer supplied, and are typically ordered with the attached storage server or device.

The FC-AL connections to specified host adapters or to other storage area network components require either 50.0/125 or 62.5/125 micron fiber optic cables. The 50.0/125 cable can be customer supplied or ordered with the SAN Data Gateway with feature # 58xx. The 62.5/125 cable must be supplied by the customer. The connector type is duplex SC. The 50.0/125 micron cables can be used for distances up to 500 m in length, while the 62.5/125 micron cable can be used for distances up to 175 m.

The SAN Data Gateway includes a CD-ROM that provides the StorWatch SAN Data Gateway Specialist, a set of management tools for the Gateway, and product documentation. These tools can be installed on a Windows NT host server or workstation that is IP network attached to one or more SAN Data Gateways, and provide an easy to use graphical interface to configure, manage, and service the Gateway. The tools enable status monitoring, and dynamic changes to the Gateway setup, providing the added convenience of centralized management of multiple Gateways. See Chapter 14, "StorWatch" on page 287 for details.

The network-based configuration requires an Ethernet network connection, and the assignment of an IP address to each SAN Data Gateway. The Ethernet connection is an RJ45 connector, and supports 10BaseT standards.

For further information about SAN Data Gateway, see the Web site `www.storage.ibm.com/ibmsan/index.htm`, and the redbook *Introduction to Storage Area Network*, SG24-5470.

9.4 IBM SAN Data Gateway Router

This section provides information related to the IBM SAN Dada Gateway Router (2108-R03).

9.4.1 Overview

The IBM SAN Data Gateway Router (2108-R03)is a SCSI to Fibre Channel protocol converter for tape libraries. It is a low-cost solution, compared to the IBM SAN Data Gateway product.

9.4.2 Key features

The IBM SAN Data Gateway Router offers:

- Industry-standard Fibre Channel host and storage area network attachment
- Investment protection of SCSI-based tape libraries
- Ease of migration between SAN connection interfaces
- A Fibre Channel port capable of up to 100 MB/s bandwidth
- Two Ultra SCSI ports (standard), either single-ended or differential
- Attachment to storage area networks via the IBM Fibre Channel Switch and Fibre Channel Storage Hub
- StorWatch SAN Data Gateway Specialist for network management of the SAN Router

9.4.3 Specifications

The SAN Router is a compact, self-contained unit. There is one standard Fibre Channel shortwave optical port for attachment to Fibre Channel-capable host systems or other SAN Fibre Channel connectivity products, and two standard Ultra SCSI ports. By selecting feature number 2820, or ordering part number 2108R3S, the two Ultra SCSI ports are

single-end SCSI. If you select feature number 2830, or order part number 2108R3D, the two Ultra SCSI ports are high voltage differential.

The two SCSI connectors are 68-pin HD P-Shell (female) connectors. SCSI cables are customer supplied, and typically ordered with the attached tape library.

The Fibre Channel port requires either a 50.0/125 or 62.5/125 micron fiber optic cable for attachment to supported host systems or SAN connection products. The cable can be customer supplied, or a 50.0/125 micron cable can be ordered with the SAN Router with feature number 58xx or by appropriate part number. The connector type is a duplex SC. The 50.0/125 micron cable can be used for distances up to 500 meters. In contrast, the 62.5/125 micron cable can be used for distances up to 175 meters.

The IBM SAN Data Gateway Router includes a CD-ROM that provides the StorWatch SAN Data Gateway Specialist, a set of management tools for the Gateway, and product documentation. These tools can be installed on a Windows NT host server or workstation that is IP network attached to one or more SAN Data Gateways, and provide an easy to use graphical interface to configure, manage, and service the Gateway. The tools enable status monitoring, and dynamic changes to the Gateway setup, providing the added convenience of centralized management of multiple Gateways. The network-based configuration requires an Ethernet network connection, and the assignment of an IP address to each SAN Data Gateway. The Ethernet connection is an RJ45 connector, and supports10BaseT standards.

Specific details on supported host systems, adapters, and SAN configurations are available on the Web at:

`http://www.ibm.com/storage/SANGateway`

9.5 IBM Fibre Channel Hub

This section provides information related to the IBM Fibre Channel Storage Hub (2103-H07).

9.5.1 Overview

The IBM 2103 Model H07 Fibre Channel Storage Hub can be used to ease the configuration of Fibre Channel Arbitrated Loops, and to add greater distances to the FC-AL topology. It is an entry-level, Windows NT supported FC Hub. The FC Storage Hub provides four short wave optical Gigabit Interface Converter (GBIC) ports, and the option to add up to three additional

long wave or short wave optical GBIC ports. The short wave GBIC ports allow fiber cable connection of up to 500 meters to the host based FC-AL initiator adapter, to the FC-AL port on the FC RAID Storage Server, or the FC port on the SAN Data Gateway or Router (2108-G07/R03). The IBM Fibre Channel Storage Hub comes in a rack mounted tray that requires 1U of rack space.

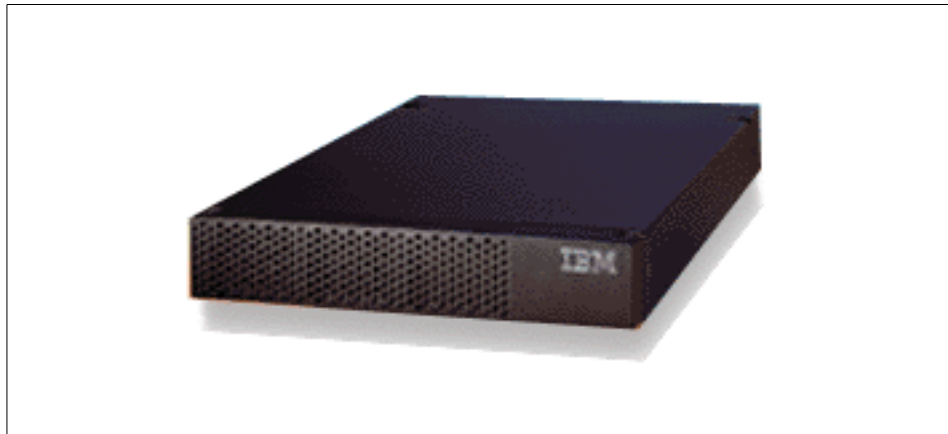A picture of the IBM Fibre Channel Hub is shown in Figure 80.

Figure 80. IBM Fibre Channel Hub

9.5.2 Key features

The IBM Fibre Channel Storage Hub offers:

- Support for high-availability configurations with multiple server and storage connections
- Interconnects multiple servers and storage systems over fiber media
- Supports data transfer speeds of up to 100 MB per second
- Enables extended distance configurations of up to 10 kilometers between servers and/or storage systems
- Provides seven hot-pluggable gigabit FC-AL ports with automatic bypass of unused ports

9.5.3 Specifications

The IBM Fibre Channel Storage Hub can be used to ease the configuration of Fibre Channel Arbitrated Loops, and to add greater distances to the FC-AL topology. The FC Storage Hub provides four short wave optical Gigabit Interface Converter (GBIC) ports, and the option to add up to three additional

long wave or short wave optical GBIC ports. The short wave GBIC ports allow fiber cable connection of up to 500 meters to the host based FC-AL initiator adapter, to the FC-AL port on the FC RAID Storage Server, or the FC port on the SAN Data Gateway or Router (2108-G07/R03). Either short wave or long wave GBIC ports can be used to connect two FC Storage Hubs, extending the distance up to an additional 500 meters with the short wave GBIC, or up to an additional 10 kilometers with the long wave GBIC. The IBM Fibre Channel Storage Hub comes in a rack mounted tray that requires 1U of rack space.

More information and specifications are available at:

`http://www.storage.ibm.com/hardsoft/products/fchub/fcshub.htm`

9.6 IBM Fibre Channel Managed Hub

This section provides information related to the IBM Fibre Channel Managed Hub (3534-1RU).

A picture of the IBM Fibre Channel Managed Hub is shown in Figure 81.

Figure 81. IBM Fibre Channel Managed Hub

9.6.1 Overview

The IBM 3534 is the Storage Area Network (SAN) Fibre Channel Managed Hub that supports entry-level work group Fibre Channel Arbitrated Loop (FC-AL) connectivity for such applications as high-availability clustering, storage consolidation, and LAN-free backup. The Managed Hub is designed for implementing entry-level homogeneous multinode server clusters and

storage systems, providing high-speed interconnections for FC-AL environments. An arbitrated loop is logically formed by connecting all eight ports on the Hub into a single loop, or the ports can be zoned into several arbitrated loops. Each port supports 100 MB/second full duplex data transfer. And with nonblocking architecture, the Managed Hub can handle simultaneous data transfer.

9.6.2 Key features

The IBM SAN Fibre Channel Managed Hub offers:

- Enables entry-level Windows NT server clustering
- LAN-free backup, storage consolidation, and remote disk mirroring
- High-speed performance utilizing nonblocking switch-based technology
- Simultaneous 100 MB/second full duplex data transfers across all ports
- Enhances availability through fault isolation
- StorWatch FC Managed Hub Specialist, a Web browser interface for configuration, management, and service
- Support of industry standard MIBs enabling standard SNMP management

9.6.3 Specifications

The IBM SAN Fibre Channel Managed Hub provides industry standard Fibre Channel Arbitrated Loop connectivity of an entry-level homogeneous cluster of host servers and storage systems. The Managed Hub has eight FC-AL ports. Seven ports support fixed short wave optical media for connecting devices on multimode fiber up to 500 meters. The eighth port is a Gigabit Interface Converter (GBIC) slot that can be configured for either short wave optical (order part number 03K9308), or long wave optical (part number 03K9307) media. Long wave singlemode fiber can be up to 10 kilometers.

The Managed Hub is an integrated rack mountable component that is serviced as a single replaceable unit. Included with the Managed Hub is a rack mounting kit, product documentation, and a power cord for connecting to the Power Distribution Unit (PDU) in the rack. The Managed Hub installs in the Netfinity RACK (9306-900) or an equivalent, industry standard 19-inch rack. A stand-alone configuration is an option. In rack-mounted configurations, 1U space and one outlet on the rack PDU are required.

Attachment to host servers and storage systems requires fiber optical cables. The short wave ports require multimode optical cables, either 50.0/125 micron optical cables for distances up to 500 meters, or 62.5/125 micron fiber

optical cables for distances up to 175 meters. If a long wave GBIC is utilized, the required fiber optical cable is a singlemode, 9.0/125 micron cable for distances up to ten kilometers.

The IBM SAN Fibre Channel Managed Hub is designed for rack mount use. More information and specifications are available at:

`http://www.storage.ibm.com/hardsoft/products/fchub/fcmhub.htm`

9.7 IBM SAN Fibre Channel Switch (Brocade)

The IBM 2109 SAN Fibre Channel Switch is used to interconnect multiple host servers with storage servers and devices, creating a storage area network (SAN).

9.7.1 Overview

The IBM SAN Fibre Channel Switch is an 8-,or 16-port Fibre Channel switch (2109-S08 and 2109-S16 respectively) used to interconnect multiple host servers with storage servers and devices, creating a storage area network (SAN). The switch allows you to build a wide range of scalable SAN solutions with high-speed, full-fabric support, and modular components. By creating an intelligent connectivity infrastructure, or Fabric, that supports connections across a wide range of host and storage types, the IBM SAN Fibre Channel Switch enables storage resources to be shared efficiently and to scale rapidly, meeting the increasing demands by users for highly available, heterogeneous access to expanding storage pools

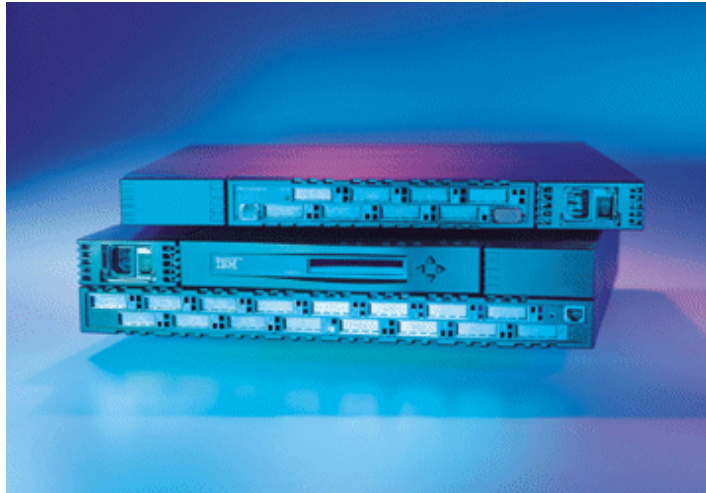Below we show a picture of both IBM FIbre Channel Switches (Figure 82).

Figure 82. IBM Fibre Channel Switches

9.7.2 Key features

The IBM SAN Fibre Channel Switch provides:

- Supports Intel-based servers running Windows NT and UNIX-based servers.

- High Performance with low latency. The Switch's non-blocking architecture provides multiple simultaneous connections, each capable of up to 100 MB/s, with maximum latency of two microseconds.

- Ports designed to support F, FL, and E-port modes of operation.

- Easy setup and configuration. Internal power-on self test and a Web browser interface usable from any Java-enabled browser on the Inter/intranet provide configuration monitoring and diagnostics.

- Scalable SAN solutions. Cascading Switches enable flexibility in configurations to accommodate many different needs, and assure high availability, and optimum performance.

- StorWatch Specialist, which allows you to control which host systems are logically connected to various storage systems and devices, providing security and access control.

- Automatic Fabric discovery. The Switch allows external host and storage systems to discover other SAN-enabled systems that are connected to the connectivity infrastructure (Fabric). This information is maintained within an internal name server database.

- Extended distance. The Switch can be configured with either shortwave or longwave Gigabit Interface Convertors (GBICs), for distances up to 10 kilometers between connections.

9.7.3 Specifications

The IBM SAN Fibre Channel Switch comes with the StorWatch SAN Fibre Channel Switch Specialist, a comprehensive set of management tools that support a Web browser interface for flexible, easy-to-use integration with existing enterprise management structures.

The IBM SAN Fibre Channel Switch has a number of selectable options. The base Switch contains four shortwave GBICs, providing four Fibre Channel port connections. You can select, in increments of one, additional shortwave or longwave GBICs up to the maximum number of ports on the selected model (eight on the S08 and 16 on the S16). A second power supply can be installed on the Switch, providing a redundant power supply. This enables dual-power source configurations to minimize power outages, and, with the hot-pluggable capability of the power supplies, allows for non-disruptive maintenance should a power supply fail. The IBM SAN Fibre Channel Switch can be used in either rack mounted or tabletop configurations. The Switch requires fiber optic cables for connection to the host systems, and storage systems or devices. These cables can be customer supplied, or ordered with the Switch.

For more information on the IBM SAN Fibre Channel Switch, visit this site:

```
http://www.ibm.com/storage/FCSwitch
```

9.8 INRANGE FC/9000 Fibre Channel Director

This section provides information related to the INRANGE FC/9000 Fibre Channel Director. The core product is also known as IBM's 2042.

9.8.1 Overview

The INRANGE FC/9000 Fibre Channel Director provides "enterprise-class" availability and Fibre Channel fabric connectivity for IBM servers and storage supporting the FICON protocol. Each FC/9000 Director can be configured with up to 64 ports, and with the High Availability Option, it provides a fully-redundant Fibre Channel fabric switch. With the FC/9000 Director Cabinet, up to two 64-port FC/9000 Directors can be configured utilizing only six square feet of floor space. The expandability and scalability of the FC/9000 Director enables migration and growth to meet your Enterprise SAN

needs. (See Figure 83). For a current list of IBM supported FICON servers and storage, refer to:

```
http://www.ibm.com/storage/director
```


Figure 83. IBM's 2042 INRANGE FC/9000 Fibre Channel Director

9.8.2 Key features

The INRANGE FC/9000 Fibre Channel Director offers:

- High performance, full duplex, 100 MB/sec per port, non-blocking switch performance.

- Industry-standard short wave laser GBICs support for up to 500 meters, and long wave laser GBICs support up to 10 kilometers.

- Scalable growth with 24 to 64 Fibre Channel ports in 8-port increments that are hot-pluggable allowing non-disruptive upgrades.

- Flexible Fibre Channel attachments include FICON, FCP Fabric, FC-AL Public and Private Loop, and cascaded Directors

- IN-VSN Enterprise Manager includes Call Home, pager, and SNMP notification if service is necessary. IN-VSN Enterprise Manager Server with private LAN attachment supports multiple Directors and up to 16 IN-VSN Enterprise Manager Clients for remote management.

9.8.3 Specifications

The INRANGE FC/9000 Fibre Channel Director provides the connectivity necessary to make enterprise information accessible and available across multiple host systems, from various systems or devices forming a Data Center Storage Area Network (SAN). Utilizing the standard Fibre Channel interconnect technology, the FC/9000 Director provides 100 MB/s transfer bandwidth on each Fibre Channel connection port. The Model 001 provides of up to sixty-four ports for high availability enterprise application servers.

The INRANGE FC/9000 Fibre Channel Director has a number of selectable options. The base FC/9000 Director contains three I/O Modules with 24 ports. You can add from one to five 8-port I/O Modules for a maximum of 64 ports. Each I/O Module can be configured with either eight short wave, eight long wave GBICs, or a mix of four short wave and four long wave, GBICs. The High Availability Option provides a second Power Module, providing a redundant power supply. This enables dual-power source configurations to minimize power outages. The High Availability Option also provides a spare Switch Module and redundant Control Module, providing automatic failover without loss of data access in case of failure. All hardware FRUs and firmware may be maintained and upgraded concurrently without loss of data access.

The FC/9000 Fibre Channel Director can be configured with either short wave or long wave GBICs for distances up to 10 kilometers. The base Director contains three FIO Modules with 24 ports. You can add from one to five additional 8-port FIO Modules (see feature number 5010) for a maximum of 64 ports. Any mixture of short wave and long wave GBICs are supported, but all ports must have GBICs ordered. GBICs must be ordered in increments of 8, either eight short wave, eight long wave or a mix of four short wave and four long wave.

For more information about the INRANGE FC/9000 Fibre Channel Director, see the Web site:

`http://www.storage.ibm.com/ibmsan/products/directors/inrange_dir.htm`

9.9 McDATA ED-5000 Enterprise Fibre Channel Director

This section provides information related to the McDATA ED-5000 Enterprise Fibre Channel Director. This unit may also be referred to as the IBM 2032 (see Figure 84).

Figure 84. IBM's 2032 McDATA Enterprise Fibre Channel Director

9.9.1 Overview

The McDATA ED-5000 Enterprise Fibre Channel Director provides a very robust and powerful connectivity solution available for implementing large Fibre Channel networks in enterprise production environments. It is designed to provide the same levels of access availability, performance availability, and data integrity achieved with IBM ESCON Directors in the world's most demanding data centers.

9.9.2 Key features

Key features for the McDATA ED-5000 include:

- Supports e-business and other mission-critical applications

- Provides enterprise-level scalability and data center-level availability

- Provides Fibre Channel connectivity for IBM S/390 and zSeries 900 servers; UNIX-based servers and Intel-based servers running Windows NT.

- Enables connectivity to IBM FICON-attached and Fibre Channel- attached storage, and to IBM UltraSCSI-attached storage with gateways and routers
- Offers 32-port Fibre Channel switch fabric with full redundancy for all active components.
- High performance, full duplex, 100 MB/sec, non-blocking switch performance.
- High Availability Option supports concurrent hardware maintenance with firmware upgrade capability and automatic failover to redundant switching, control, and power supply modules in case of failure.
- Uses Enterprise Fabric Connectivity (EFC) Management software to centralize control for multiple Enterprise Fibre Channel Directors across the enterprise.

9.9.3 Specifications

The Director provides a wide range of configuration options. One port card provides four Fibre Channel ports, with each port delivering 100 MB/sec, full-duplex data transfer. The ED-5000 Director supports full nonblocking 32-port switch performance.

Other configuration options include short wave laser cards (transmission distance up to 500 meters), long wave laser cards (transmission distance up to 10 kilometers) or combination cards (1 long wave port and 3 short wave ports). You can combine cards for a maximum of thirty-two ports.

The McDATA ED-5000 includes software to easily manage your SAN environment. Enterprise Fabric Connectivity (EFC) Management Software. EFC Management software centralizes the management of multiple, distributed Directors in an enterprise-wide Fibre Channel fabric. Designed for usability, this flexible software simplifies the addition of managed Directors as the enterprise Fibre Channel fabric grows. Leveraging state-of-the-art Java technology, operators can use the EFC Management software remotely from anywhere in the enterprise.

EFC Management software provides a scalable, modular architecture consisting of the EFC Server, the EFC Manager application, and the EFC Product Manager application. The architecture is implemented through a multi-tiered client/server networking model. You can access all management, monitoring, and diagnostic functions by running the EFC Management client application on a workstation located anywhere in the enterprise.

McDATA ED-5000 is normally mounted in the McDATA FC-512 Cabinet. One or two Directors may be mounted in the McDATA FC-512 Cabinet, a single 24" wide cabinet with a footprint of only 7 square feet. This cabinet includes the Ethernet hub and space for installing a EFC Server. Dual power cords support the HA features. The FC-512 Cabinet provides 36U space for mounting the Directors. It also includes the Ethernet hub and 3U space for installing an EFC Server. Dual power cords support the High Availability features.

For more information about McDATA Enterprise Fibre Channel Director, visit this site:

`www.mcdata.com/efcdirector`

Chapter 10. Other storage solutions

This chapter addresses other storage solutions that should be considered for your storage configuration.

10.1 IBM 3466 Network Storage Manager

The IBM 3466 Network Storage Manager, an IBM Storage Smart Solution, is an integrated, pre-configured system for complete network backup and archive support.

10.1.1 Overview

The IBM 3466 Network Storage Manager is a member of the Seascape family offering a complete integrated hardware solution and software functions that provide automated backup and recovery, advanced storage management, and archive and disaster recovery of business data stored in a multi-platform network environment. The 3466-C01 includes powerful RISC processors, increased memory, high-speed Serial Storage Architecture (SSA) disks, and customer-selectable, attached tape environments. The complete, integrated solution supports more than 30 server and workstation platforms and can be connected to enterprise networks through FDDI, Ethernet, ATM, and Token Ring interfaces. The 3466-C01 is Fibre Channel capable and allows attachment to your IBM 2108-G07 SAN Data Gateway (SDG) for library attach and sharing, as well as other SAN functions. The 3466 will still support direct tape library attachment.

The IBM 3466 Network Storage Manager Software Package Version 4 (5765-E66), pre-loaded and configured, includes TSM Version 4.1 (Tivoli) clients. Available optional features include: Managed System for SAN providing LAN-free data transfer, Disaster Recovery Manager, Space Management, Managed Library support for automated tape libraries, and Tape Library Sharing. Open Systems support for AFS and Distributed File System (DFS), and Server-to-Server Communication features are now included in the basic support provided in the pre-loaded Software Package. (Figure 85 illustrates these NSM components.)
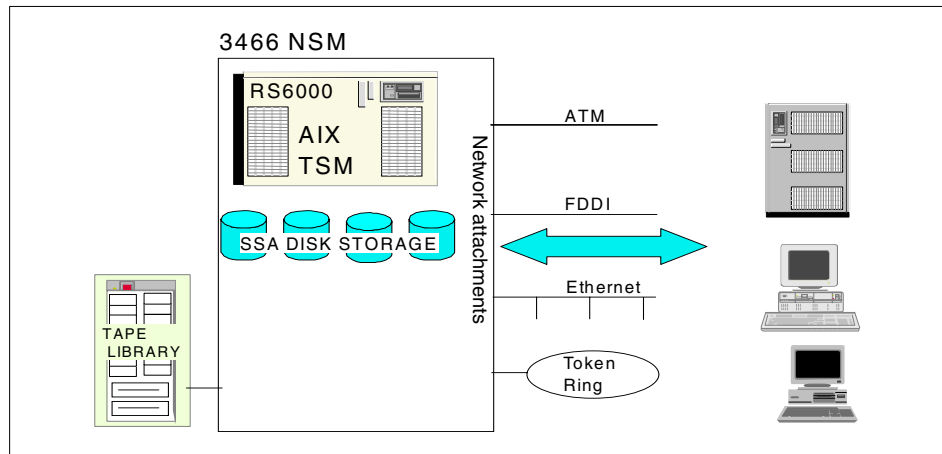
205

Figure 85. IBM 3466 NSM components

10.1.2 Key features

The 3466 NSM includes the following key features:

- Complete and integrated data storage solution

  This feature provides an easy-to-implement integrated hardware and software solution for managing backup, recovery, and archive of network distributed data.

- Rich functionality and multiplatform support

  These features include Tivoli Storage Management software for storage management, backup, archive, and disaster recovery of all major IBM and non-IBM multiplatform, distributed data.

- Two basic disk cache configurations

  - JBOD
  - RAID-5

- Upgrades to grow with your needs

  The 3466 NSM is available in a range of capacities and technologies, to best suit the needs of your network computing environment:

  - Disk capacities from 72 GB to 864 GB
  - Increased processor power (up to 4-way 340 MHz RISC Processor)
  - Up to 4 GB of internal memory
  - 1 to 3 additional network interfaces

- 3466 Remote Support Facility

  Provides the basis for rapid diagnostic and remote service capabilities.

  Note: Each Remote Support Facility will require its own dedicated analog (customer-supplied) telephone line for installation.

- Disaster recovery multiple server support

  Network Storage Manager can be configured so the machine hardware can be used by multiple Tivoli Storage Management servers, for the purpose of hot standby only, in disaster recovery mode. This is a customer-implemented option which can be activated by following instruction scripts furnished by IBM.

10.1.3 Specifications

The 3466 NSM manages client data. The clients can be desktop PCs, workstations, file servers, database servers, and so on. Client data files are transmitted by way of the network to the 3466 and written to storage. The storage of the data files can be on disk or on tape.

The Tivoli Storage Manager server provides centrally administered storage management services in the distributed environment. These services are based on a set of defined policies or rules, and you can automate and schedule many of the services to make the best use of your resources.

The IBM 3466 Network Storage Manager Model C01 consists of:

- A 2-Way SMP RISC processor with 340MHz clock and 1 GB DRAM
- Control console
- One Dual SCSI adapter card
- One SSA adapter card
- One (minimum) to four (optional) communication adapter card(s)
- Two internal 18.2 GB hard drives
- IBM 3466 Network Storage Manager Software Package, (5765-E66), pre-loaded (separate order required)
- One to three IBM SSA disk enclosures containing up to sixteen 18.2 GB disk drive modules each, totaling up to 864 GB disk storage capacity.
- No tape library storage.

The Model C01 can use one or more of the following libraries:

- IBM Magstar MP 3575 Tape Library with 3570 Drives

- IBM Magstar 3494 Tape Library with 3590 Drives

- IBM 3584 Ultra-Scalable Tape Library with 3580 Linear Tape Open (LTO) Ultrium Tape Drives

- Network Storage Manager supported tape library supplied by the customer

10.2 Network Attached Storage (NAS)

The IBM Network-Attached Storage (NAS) is an innovative new storage solution that increases throughput and simplifies storage administration, maintenance, and management. NAS uses a LAN interface (Ethernet) to connect to a system, enabling virtually transparent access and file sharing across NFS and CIFS environments.

10.2.1 Overview

A NAS solution is basically made up of one or more very specialized servers that provide storage services through the LAN. In NAS there is also a network connecting the storage devices and the servers. But in this architecture, the network used to accommodate the storage data flow over a conventional LAN. There is no need to build an entire new network to implement a NAS solution; the existing network could be used. If required, a dedicated LAN could be built just for the storage data flow, depending on the performance level requirements.

By making storage devices LAN addressable, users on the LAN, running any operating system, can address the NAS storage by means of a common access protocol, for example, Network File System (NFS). In addition, a task such as backup to tape can be performed across the LAN, thus enabling sharing. (Refer to the redbook, *Introduction to Storage Networking*, SG24-6137.)

The IBM 8658 xSeries 150 (Figure 86), a Point Solution server, is an advanced network attached storage (NAS) appliance for today's e-business. This platform design supports multiple file protocols enabling Windows, Novell, and UNIX clients to access storage. It delivers excellent value, combining X-architecture integration, leading systems management capabilities, and powerful Windows 2000 technology.

Figure 86. IBM 8658 x-series 150 Network Attached Storage

Two models are offered:

- Model 8658-33Y

  The IBM xSeries 150 Model 33Y has an 800 MHz processor, 256 MB of memory, ServeRAID-4L, two LAN adapters, and 3 x 36GB hard drives in a tower platform.

- Model 8658-3XY

  The IBM xSeries 150 Model 3XY has two 800 MHz processors, 1 GB of memory, ServeRAID-4H, four LAN adapters, and 6 x 36GB hard drives in a rack platform.

10.2.2 Key features

xSeries 150, using the popular xSeries 230 platform, is preconfigured from the factory with hardware features and software designed, to handle NAS applications and general purpose networking applications in heterogeneous (IBM and non-IBM servers) and non-heterogeneous (IBM servers) network environments.

- A fully integrated storage solution
- Redundant hot-swappable components for high availability

- Simple LAN installation through a Web browser interface

- Open system design that enables easy integration with the existing network, as well as a smooth migration path as storage needs grow

- Support multiple file protocol (NFS, HTTP, CIFS, and NetWare)

- Scalable from 108 GB to 1.7TB

10.2.3 Specifications

Workgroup model -33Y:

- Tower-based configuration
- 800 MHz Pentium III processor
- 256 MB system memory
- Two 10/100 Mbps Ethernet controllers
- 109.2 GB of storage — expansion to 218.4 GB

Departmental model -3XY:

- Rack-based configuration
- Two 800 MHz Pentium III processors
- 1 GB system memory
- ServeRAID-4 H Ultra160 controller
- Four 10/100 Mbps Ethernet controllers
- 218.2 GB of storage - expansion to 1.74 TB with EXP300 Storage Units

Common features:

- Preloaded and optimally tuned with Windows Powered OS and appliance software stack
- Supports multiple file protocols
- Two 250-watt, hot-swap, redundant, auto-ranging power supplies
- Integrated Advanced System Management Processor

Part 3. Understanding storage management software

Chapter 11. Introduction to storage management software

In this chapter, we describe the role of storage management software and some of its capabilities. We then define technologies, standards, and building blocks that are used in storage management software, and finally introduce IBMs storage management software.

Storage administrators use many different software tools while performing their jobs, for example, graphic software or spreadsheets and financial planning programs. In this section, we concentrate only on software that is specialized to perform storage management tasks.

11.1 Roles of storage management software

In Chapter 1, "Enterprise Storage Management overview" on page 3, we discussed the challenges currently facing storage administrators. Today's storage management software was developed to meet current challenges posed to storage administrators. As the complexity of your storage environment increases, storage administrators must become more dependent on automated solutions for enterprise storage management. It will become increasingly difficult to try to manually manage environments.

In today's OS/390 environment, storage administration tasks are highly automated, and these facilities must extend into the distributed storage environment. The purchase or development of software to manage both storage devices and the data contained on them will continue to be justified by the ability of this software to automate tasks that would otherwise have been performed by administrators.

Roles of storage management software include:

- **Implementing and enforcing policies** — Although storage administrators define policies, the storage management software can be used to implement and enforce these policies.

- **Performing repetitive actions** — For example, data movement or performing backups.

- **Monitoring the usage and performance** of storage resources.

- **Producing reports** — Both on how these products themselves are performing, and also on how other storage resources are performing.

- **Identifying problems** with the storage environment by reporting on deviations from acceptable thresholds for both hardware and software.

213

Currently, the major role of storage software is in automating tasks that would otherwise have required an administrator to perform. Automating task provides several advantages:

- **Freeing storage administrators to perform other tasks**. Storage administrators are a highly skilled resource. Allowing automation to perform basic repetitive tasks frees storage administrators to perform more complex work.

- **Reducing the chance of errors**. Performing repetitive manual tasks inevitable leads to errors, software is less subject to distraction than humans.

- **Performing tasks 7x24 with no overtime required**. Software does not require sleep.

- **Speeding the response to problems**. By implementing thresholds, software can potentially generate alerts immediately before actual problems have occurred, without relying on someone to notice that something is wrong and act on this observation.

Storage management software has, by no means, removed the requirement for storage administrators. However, it does allow your administrators to spend their time on tasks requiring skills that are unavailable in today's storage management software; for example:

- **Specific problem resolution**. Software can identify problems based on threshold exceptions, and may be able to implement a bypass to the problem. However, in some cases it still requires human intervention to determine the root cause of the problem and correct it, rather than just working around the problem's symptoms.

- **Physical layout planning**. Some software will, for example, assist in physical layout planning, or by running simulations of the effectiveness of planned changes to the current environment (SMS Optimizer). However, planning in the storage environment still requires people to make the final decisions.

- **Implementing new policies**. Developing and implementing new policies still requires human skills and familiarity with the storage environment, although storage administrators will certainly use data generated by storage management software to provide supporting data for the decisions that are made.

The following redbooks help you implement new policies in the OS/390 environment:

- *Get DFSMS FIT: Fast Implementation Techniques*, SG24-2568

- *DFSMS FIT: Fast Implementation Techniques Process Guide*, SG24-4478

- *DFSMS FIT: Fast Implementation Techniques Forms and Foils,* SG24-2570

- *DFSMS FIT: Fast Implementation Techniques Installation Examples*, SG24-2569

11.2 Introduction to standards, technologies, and building blocks

The following are standards, technologies, and building blocks used in storage management software. We have included these items to provide a level-set of terminology. Items are included from both the mainframe and distributed storage environments.

11.2.1 X/Open Data Storage Management API

X/Open data storage management (XDSM) API is a standard defined by the Open Group. It defines application programming interface for programs to use functions provided by storage management software. Tivoli Storage Manager is an XDSM compliant product.

11.2.2 X/Open Backup Services API

X/Open Backup services API (XBSA) is a standard defined by the Open Group. It defines application programming interface for programs to use functions provided by storage management software.

Tivoli Storage Manager is an XBSA compliant product. For more details, see:

```
http://www.storage.ibm.com/storage/software/adsm
```

11.2.3 Management class

Management classes define a a set of actions that storage management software performs on data. Management classes are used to determine factors such as:

- Backup frequency
- Backup retention
- Archive retention
- Retention
- Class transition
- Data movement

Both Tivoli Storage Manager and DFSMS/MVS use structures called management classes to manage data, although the implementation of management class differs in each product. Within DFSMS a management class is defined by the DFSMSdfp component and is used by this and other components, for example DFSMShsm and DFSMSrmm.

11.2.4 Filtering

Filtering is a technique that allows pattern matching for data. Only data that matches the pattern set by the filter is allowed to pass through the filter. Filtering is most often used to specifically target objects for processing, either by inclusion or exclusion. Often filtering techniques also allow the specification of boolean operators to change the scope of a filter. The major advantage of filtering is that changes to filters need to be made less frequently than changes to explicit specifications. To be used effectively, filters require standardization of specification of the resource used for filtering.

11.2.5 Volume pooling

Volume pooling enables multiple physical storage resources to be one logical storage resource; thus, many physical volumes, the pool, appear as one logical target.

AIX provides logical volume manager (LVM) to do this. Other selected UNIX operating systems also provide volume pooling capability.

Tivoli Storage Manager pools its own storage resources, such as backup volumes, archive volumes, and migration volumes.

DFSMS/MVS provides the storage group (SG) to do this. Storage groups can apply to magnetic disks, optical disks, and tapes. All volumes in a storage group are potentially available to satisfy an allocation directed to any individual volume. DFSMShsm pools its own storage resources, such as migration, backup, and dump volumes. DFSMSrmm enables pooling of tape volumes as well as allowing allocations to be directed to specific volume pools.

11.2.6 Hierarchical Storage Management

The value of information frequently diminishes over time, and the likelihood of an immediate demand for data falls off dramatically over the month or quarter to which the data relates. The longer that data has been unreferenced, the less likely it is to be reaccessed. The use of other data may be strictly periodical; for example, it may only be required for the first two days of the

month. In both these cases, having the data remain on expensive DASD storage, when it is not in use, may not be the best use of the DASD resource. Installations usually have a mixture of different devices where data can reside; these devices offer a range of different performance and availability characteristics. Together, these devices form a storage hierarchy.

An excellent way to balance the requirement for expensive disk storage with data usage patterns is to study the life cycle of the data, and manage to this. The first consideration must still be data integrity. Fast, inaccurate data is worse than no data at all, but consider how quickly you need access to the information. Data used every day, or information your staff uses while dealing with customers, must be available on demand; data used only to create quarterly summary reports may have controlled access with planned availability.

Hierarchical storage management (HSM) moves data to the most appropriate level within the storage hierarchy. Generally, this involves moving data from one storage media to another as its pattern of use dictates; and optimizing the use of your most expensive media, usually the fastest, by ensuring that the data occupying this resource is producing the best return on investment.

Both Tivoli Storage Manager and DFSMShsm provide HSM function. We discuss this function in detail in the following sections 12.3, "Space management" on page 233, which discusses HSM functions within Tivoli Storage Manager; and 13.3, "Space management" on page 260, which discusses HSM functions within DFSMS.

HSM can also be seen in products such as DFSMSrmm, where data may be off-loaded from data servers according to a management class to preserve expensive space in the server for scratch mounts.

VTS also employs the HSM concept of providing internal hierarchies of storage resources to optimize data placement, although there is no external control of this. See 8.4, "Virtual Tape Server" on page 166 for details.

A component of HSM is sometimes called *space management*, although the functions traditionally assigned to space management are more limited than those covered by true HSM. Space management is limited to those functions dedicated to ensuring that sufficient space exists in certain data pools.

Tape mount management (TMM) is another implementation of HSM (see 13.8.2, "Tape Mount Management" on page 279).

11.2.7 Backup

A backup is a copy of data. The copy may be identical in format to the original, but stored on a different media or with a different name, or the copy may be stored in a different format than the original, but which can be returned, restored, to the original format if required. There are a number of different types of backup that may be taken.

11.2.7.1 Full backup

Full backups are backups that include all the data from a specified source. They are usually based at the volume level and include both changed and unchanged data. They may also include representations of unallocated or allocated but unused space.

11.2.7.2 Incremental backup

Incremental backup processes only those files that have changed since the backup process last ran, and optionally, those files that have been newly allocated, whether or not they are marked as having been changed. An incremental backup processes only a subset of the data that would be processed during a full backup. Incremental backups usually require that a changed or altered flag is set by updates of the data. The incremental backup process generally turns off this flag.

11.2.7.3 Differential backup

A differential backup backs up those files that have changed since a full backup was taken, regardless of whether they have been processed by an incremental backup. An example is a backup of all data with the date last changed or date allocated greater than a specific date, regardless of whether the changed flag is on or off.

11.2.7.4 Physical backup

Physical backups process data at the track level and contain device dependent information. These backups can only be restored to a device type of the same geometry as the source volume and with a capacity the same or greater than the original volume. You may be able to partially restore dumps to smaller volumes. Physical dumps can process either at the volume or data set level. Stand-alone restore programs can often only process data from physical dumps.

11.2.7.5 Logical backup

Logical data set backup processes data sets as logical entities. Because the environment at a remote site may differ from your environment, this is

recommended for disaster backup, as it allows data to be restored to unlike devices.

Logical backup allows you to process either all the data or a volume or data set. You must use this type of backup if you need the ability to restore the data to an unlike device type.

11.2.8 Availability management

Availability management is the process of ensuring that appropriate backups exist to allow for recovery from incidental computer system outages, for example a disk or tape failure or a user error. Techniques that may be used include mirroring data or performing regular backups. Availability management must provide the ability to recover data in a timely manner and to an acceptable level of currency from a number of different types and severity of errors.

11.2.9 Disaster recovery

Disaster recovery is a specialized case of availability management designed to protect against the loss of data in a disaster. Essentially, a disaster can be described as a problem that cannot be resolved on site within an acceptable period of time, and that requires the use of an alternative computing facility. Obvious examples are major fires, explosions, evacuations, and so on. Sometimes, however, a minor problem that appears to be recoverable through normal problem management procedures can also escalate into a disaster as a succession of small problems interact with each other.

Management must determine the time frame that moves an outage from problem status to disaster status. Most organizations will accomplish this by performing a business impact analysis to determine the maximum allowable downtime for critical business functions.

The premise of disaster recovery is that you have copies of vital data sets stored at a location separate from your computer site, and that a computer system of approximately the same capabilities is available to resume operations. Since the data must be stored at another site, either it must be on a portable media (tape) or you must automatically transmit it either by mirroring data or using network transfer.

Disaster backups are made specifically for recovering data and applications following a disaster. You should not rely on regular backups for disaster recovery, since disaster backups require special considerations and handling. They are stored off-site and are recovered at a remote site. How remote the recovery site is depends upon the type of disaster (in case of fire, the

recovery site can be around the corner; in case of earthquake or flood, it should be many miles away).

Disaster recovery data should include all data sets critical to your operation, such as critical application data sets, RACF, inventory data sets, system data sets, and catalogs.

11.2.10 Compression

Compression is a means of reducing the size of data either transmitted or stored. It is usually accomplished by removing repeated characters. However, there are many different compression algorithms that have been patented, and many of these are in use in storage subsystems and software. For example, DFSMS uses compression in different environments.

- DFSMSdss has the ability to compress data during data set dumps.
- DFSMShsm can compress data during migration.
- DFSMSdfp can compress data before it is written to disk, using a standard compression dictionary; or it can create tailored compression dictionaries for some data.

Hardware can also have the inherent ability to compress data — for example, 9672 processors and IDRC on tape units, or RVA DASD.

If a storage subsystem has its own compression, storage software omits it and passes data to the subsystem so that installations can save processor resources.

11.2.11 Inventory management

Inventory management is the process of categorizing objects and ensuring that they are being used to the maximum profitable extent. For example, DFSMShsm manages a catalog of tape volumes, some of which are full, while others may be empty. DFSMShsm is able to move data to maximize the utilization of the tapes that it owns, allowing it to release tapes that it has removed data from. Similarly, DFSMSrmm manages your total inventory of tape volumes, ensuring that volumes containing valid data are not overwritten, that data no longer required is discarded, and that data needing to be moved to other locations is identified.

11.2.12 Common user interface

The common user interface allows storage administrators to manage storage resources easily.

The Tivoli Storage Manager backup-archive client has a *command-line interface* (CLI) and a GUI, and a Web-based interface that allows remote graphical management.

The StorWatch product family is designed with a consistent look-and-feel for each of the products which have browser-based user interfaces.

DFSMS/MVS provides an Interactive Storage Management Facility (ISMF).

11.2.13 Stacking

Stacking places multiple files or data sets onto a single storage resource such as tape. Stacking is important because the latest tape technologies have greatly increased the capacity of tape volumes. Stacking of related data on a single tape (for example, Tivoli Storage Manager's collocation) can also reduce the number of tape mounts that may be required to retrieve data.

11.2.13.1 Software data stacking

The following are some software data stacking solutions:

- Both DFSMShsm and Tivoli Storage Manager fill tape media by stacking many logical files into one physical data set.
- JCL use of the LABEL=x parameter allows manual stacking of data sets on a tape.
- DFSMShsm is able to stack multiple dumps on one output tape using the DEFINE DUMPCLASS(STACK(nn)) command.

11.2.13.2 Hardware data stacking

Examples of hardware that provide data stacking are these:

- IBM VTS, where multiple virtual tape volumes are located on one physical tape volume.
- The MVS VTOC, which allows multiple data sets to reside on one MVS volume.

Chapter 12. Tivoli Storage Manager

In this chapter we describe Tivoli Storage Manager, which is the new name for the well-known IBM software, ADSM. Tivoli Storage Manager is software that manages enterprise-wide data from a central point.

Reading through this topic, you will discover many functions in the Tivoli Storage Manager solution that are similar to those in DFSMS/MVS, as explained in Chapter 13, "Data Facility Storage Management Subsystem" on page 253. This could be a good opportunity to consider consolidating your storage management teams from both worlds, open systems and S/390, for a unique and more efficient storage management team. See 16.1, "Understanding enterprise storage administration" on page 305 for details.

12.1 Overview

Tivoli Storage Manager is an IBM client-server solution for distributed storage management; it supports a wide variety of IBM and non-IBM platforms for both small and large systems.

Storage is managed between the clients and one or more Tivoli Storage Manager servers, through the network. The server stores the client's data on disk, tape, or optical, using a local database (see Figure 87). The Tivoli Storage Manager database points to the client's data in storage and takes part in the data recovery.
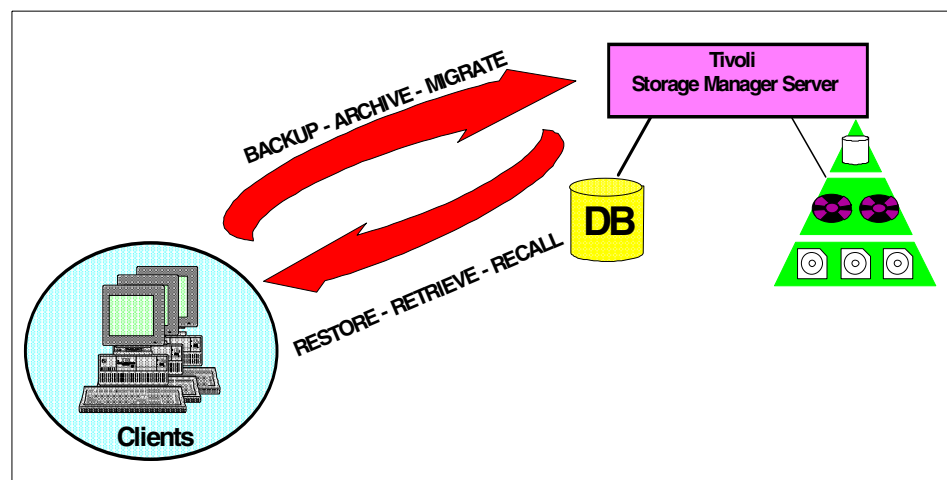
Figure 87. Tivoli Storage Manager overview

223

12.1.1 Key features

Tivoli Storage Manager includes the following key features:

- **Availability management**, to back up and archive data, including protection in case of disaster

- **Space management**, to migrate data from the client to the server

- **Policy control**, to allow different types of data to be treated differently

- **User interface**, to access the client functions and administer the server

- **Scheduling**, to centralize and automate the client's data management from the server

Other specific features are the multiplatform environment; the Tivoli Data Protection modules which back up some enhanced databases; and the communication capability among several Tivoli Storage Manager servers.

See *Tivoli Storage Manager Version 3 Technical Guide*, SG24-5477, and *ADSM V3 MVS Administration Guide*, GC35-0277, for details.

12.1.2 Product components

Tivoli Storage Manager has two major components, as shown in Figure 88.
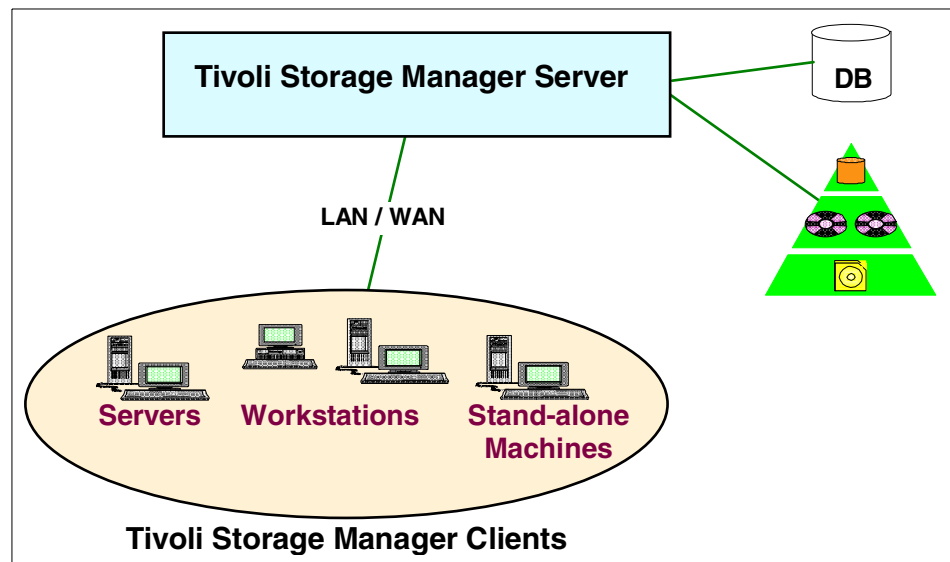
Figure 88. Tivoli Storage Manager product components

- **Tivoli Storage Manager client** — The Tivoli Storage Manager client can be a personal computer, a workstation, or a file server. It contains the Tivoli Storage Manager client code and holds the critical data that has to be protected. The client code consists of the backup and archive client (Install Archive and Rapid Recovery), the HSM migration client (see 12.4.3.3, "Hierarchical Storage Management" on page 240), and the graphical and command line user interfaces to manage the client (see 12.7.5, "Multiplatform support" on page 248).

  The client's data is sent to the Tivoli Storage Manager server through the network, through a communication protocol.

- **Tivoli Storage Manager server** — The server stores the data on its disk or tape storage pools and uses a database and recovery log to track the location of the data in the storage pools (see 12.7.5, "Multiplatform support" on page 248). The Tivoli Storage Manager database and the recovery log should be mirrored. It also provides the following software:

  - **Server-to-server:** To use the TARGET server as device class for Primary Storage Pool for the SOURCE server, exploiting the virtual volume function.

  - **Enterprise administration:** To manage an enterprise of Tivoli Storage Manager servers through a central administration.

  For added licenses, the customer will need:

  - **Network enabler:** To allow communication between the server and clients or other Tivoli Storage Manager servers through TCP/IP, APPC, Netbios and IPX/SPX. This license is not applied in the case of MVS-OS390, VM, and AS400 platforms.

  - **Extended device support:** Required for use large capacity storage devices, such as IBM 3494.

  - **Open system:** To allow backup/restore of the following type of data on Tivoli Storage Manager server:

    • S/390 UNIX Client

    • AIX AFS/DFS client

  - **User registration:** The number of clients that use, through a communication protocol, the services of the Tivoli Storage Manager.

  - **Data Protection Application:** Used to back up databases (see 12.7.5, "Multiplatform support" on page 248).

  The server code contains an *HTML page*, and thus can be administered from all the Web browser workstations (see 12.5, "User interfaces" on page 242).

12.1.3 Operating environment

As a client-server application, Tivoli Storage Manager supports a variety of both IBM and non-IBM platforms, and commonly used communication protocols (see Figure 89).
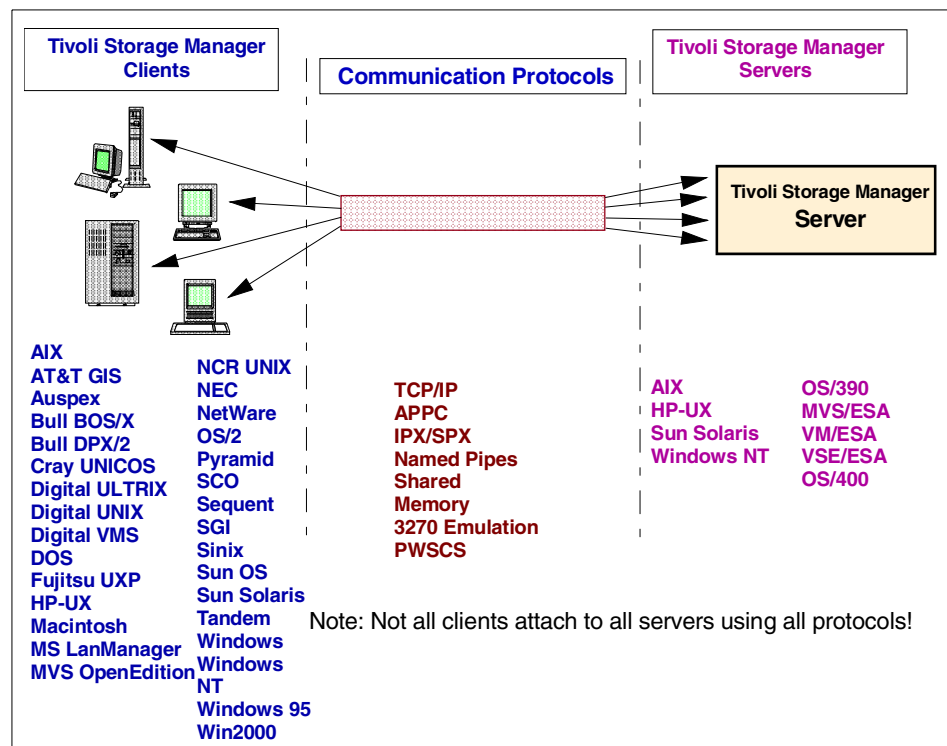
Figure 89. Tivoli Storage Manager operating environment

For details, see 12.7.5, "Multiplatform support" on page 248, *Getting Started with ADSM,* SG24-5413, or go to the Web site:

```
www.tivoli.com/storage/adsm/index.htm.
```

12.2 Availability management

In this section we describe the Tivoli Storage Manager availability management functions that secure the client's data.

12.2.1 Backup and Restore

Backup and restore are basic functions that provide fast data protection (see Figure 87 on page 223).

12.2.1.1 Backup

Backup enables you to save copies of files from the Tivoli Storage Manager client to the Tivoli Storage Manager server. Thus, your data at the Tivoli Storage Manager client is protected in the event of data lost due to a hardware or software failure, accidental deletion, or logical corruption. You can back up individual files and directories, and you can exclude specific files from the backup.

Tivoli Storage Manager has two different backup functions:

- **Incremental backup** backs up only those files that have changed. The first time you back up, you do a full backup; the next time you back up, only the changed files are sent. Incremental backup reduces the amount of data sent daily and has a major impact on both network utilization and performance.

- **Selective backup** enables you to explicitly back up a few files, whether or not they have changed.

The number of versions of backup files to keep is defined in Tivoli Storage Manager's management class (see 12.4.3, "Management classes and client options files" on page 239).

You can also activate the Tivoli Storage Manager software compression feature, which compresses the data on the client during the backup and decompresses the data also on the client during the restore. Regardless of whether or not you activate the Tivoli Storage Manager software compression feature, your storage devices can activate the hardware compression. Tivoli Storage Manager's compression feature charges the client CPU but improves the network. Thus, its activation depends on your environment.

12.2.1.2 Restore

You can restore files or directories that have been backed up.

You can request that Tivoli Storage Manager:

- Restore the current version or a specific backup version of the files.
- Restore the files to a directory other than the original directory.
- Restore the files to a point-in-time. You can restore the files to a particular date and time, for instance, July the 17th at 06:00pm.
- Replace the files (or not) if they still exist on the client machine.

Tivoli Storage Manager enables you to do a cross-client restore, within an operating system family. For instance, you can restore files saved on a Sun client to an AIX client.

12.2.2 Archive and retrieve

Archive and retrieve functions enable you to save copies of files for some period of time.

12.2.2.1 Archive

You use archive when you want to place copies of files in long-term storage (whether for legal reasons, such as vital records, or for auditing purposes).

Archive copies are never replaced with more current versions as backups are. They are preserved exactly as you store them, and for as long as you specify retention at the archiving request. After the files have been archived, you can keep or delete them from the Tivoli Storage Manager client.

When you archive files, you can associate a description with them. Therefore, when you retrieve the files, you can easily identify all files that match a given description (see Figure 90). The default description is *Archive Date: mm/dd/yyyy Time hh:mm:ss*, but you can modify it.

The number of days to keep archived files is defined in Tivoli Storage Manager's management class (see 12.4.3, "Management classes and client options files" on page 239).
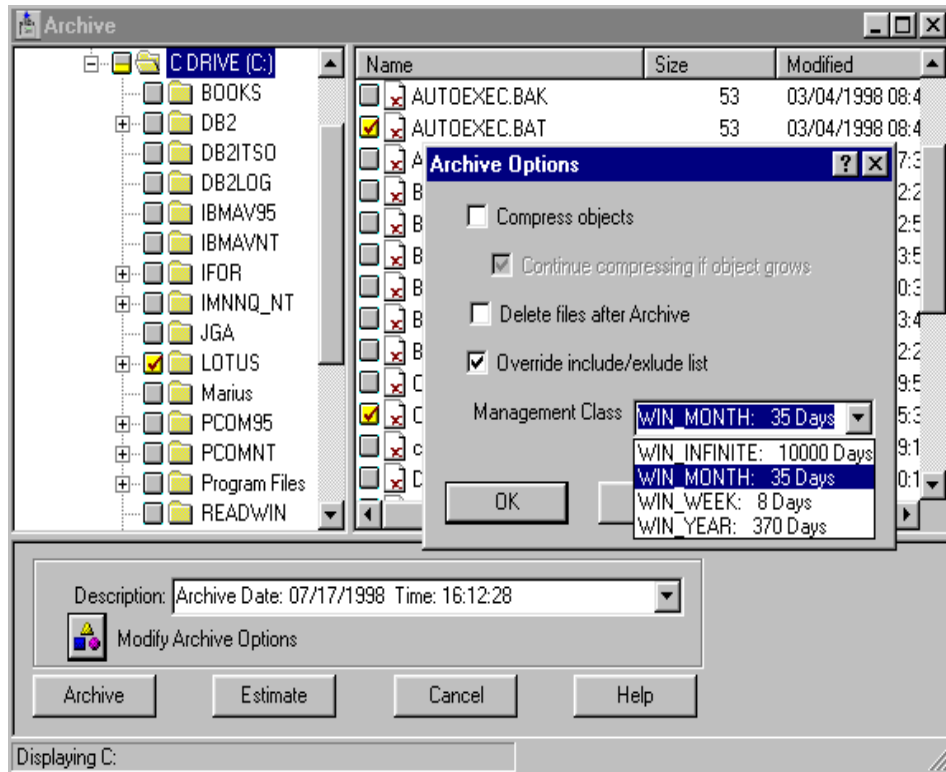
Figure 90. Tivoli Storage Manager Archive window

12.2.2.2 Retrieve

The Tivoli Storage Manager Graphical User Interface (GUI) retrieve function displays archived files grouped by their archive descriptions. Expanding the description tree displays the individual directories and files of the archive package.

Figure 91 shows the **Retrieve** window, where a list of packages is displayed (listed by archive description). One of the packages is expanded, and one file is selected for retrieve.
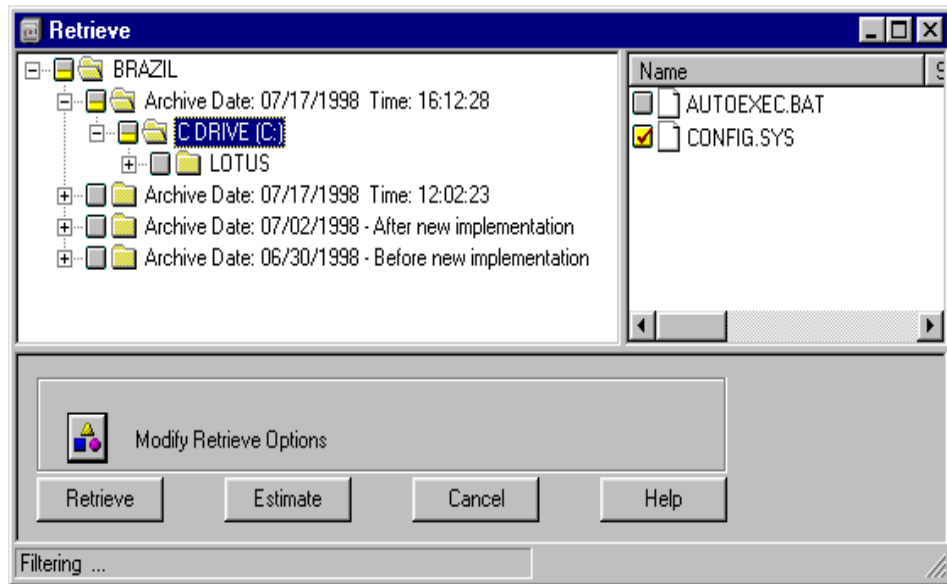
Figure 91. Tivoli Storage Manager Retrieve window

12.2.3 Instant Archive and Rapid Recovery

Instant Archive is a new function that allows the server to create a snapshot copy of client's latest backup from the Tivoli Storage Manager server storage onto a portable media. A new term, *Backup Sets,* is introduced with Instant Archive.

What is a Backup Set?

- A collection of files backed-up from one client
- Stored and managed as a single object
- On specific media or server storage
- Containing client node's active, backed up files

Note: The Backup Set is not a file system image.

The important enhancement over a normal backup or archive is that you can bring this portable media (tape, CDROM, or FTP) to the location and then do the restore there, without requiring a Tivoli Storage Manager server. This allows *Rapid Recovery* of the client when there is no access to a Tivoli Storage Manager server.

12.2.4 Tivoli Disaster Recovery Manager

A disaster is a catastrophic interruption of business processing that destroys the Tivoli Storage Manager server, or Tivoli Storage Manager clients, or both.

Tivoli Disaster Recovery Manager (DRM) is an optional Tivoli Storage Manager feature. It facilitates recovery of business applications from backups stored off site, plus the capability to use server-to-server virtual volumes for storage pool and database backups.

DRM provides:

- An automated and auditable recovery plan for the Tivoli Storage Manager server.
- Offsite recovery media management.

  You determine which volumes have to be moved from an offsite to an onsite location for recovery purposes.
- Storage of client recovery information in one place.

  You store client machine information, such as machine location and software and hardware characteristics, in the Tivoli Storage Manager database.

Disaster recovery is accomplished with Tivoli Storage Manager as follows (see Figure 92):

1. Back up client data to the Tivoli Storage Manager server.
2. Save the client recovery information in the Tivoli Storage Manager database.
3. Back up the storage pools and store the media offsite.
4. Back up the server database to removable media and store the media offsite.
5. Generate the disaster recovery plan file to assist with Tivoli Storage Manager server recovery.
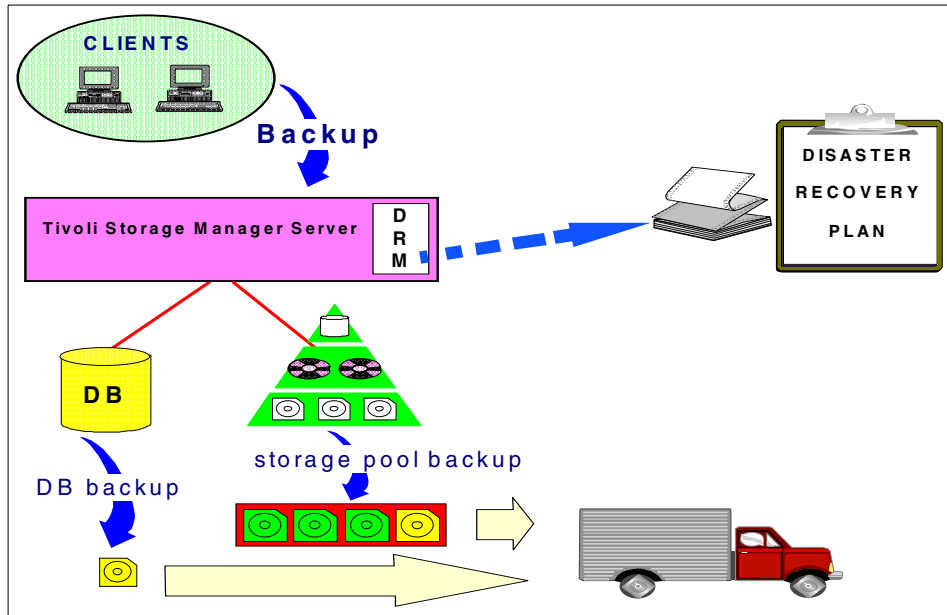
Figure 92. Tivoli Storage Manager Disaster Recovery Management

Recovery can be performed at an alternative site, on replacement hardware, by people unfamiliar with the backup applications.

After a disaster, you need:

- The disaster recovery plan file, the backup volumes, and the Tivoli Storage Manager database backup

- A suitable replacement machine for the Tivoli Storage Manager server. Use the recovery plan file to rebuild the Tivoli Storage Manager server.

- A suitable replacement machine for the Tivoli Storage Manager client. DRM gives you the client information that you stored in the database.

After you have installed the hardware, follow these steps:

1. Restore the operating system software.
2. Restore the communication protocol.
3. Restore the Tivoli Storage Manager client code.
4. Restore the Tivoli Storage Manager client options files.
5. Restore the client file systems from the primary storage pools.

For more information about recovering a Tivoli Storage Manager client, see: *ADSM for MVS DRM Administrator's Guide and Reference*, GC35-0238, or *ADSM Disaster Recovery Manager: Stand-alone Disk Image Dump and Restore Guide*, GC35-0343.

12.3 Space management

The space management function is provided through an optional product called Tivoli Space Manager. It gives clients access to the storage devices of the TSM server, thus extending their local disk capacities (see Figure 93). It uses the Hierarchical Storage Management (HSM) technique to migrate rarely accessed files automatically and transparently to TSM storage, freeing administrators and users from manually performing file system pruning tasks, and ensuring that sufficient free storage is available in the workstation or server.
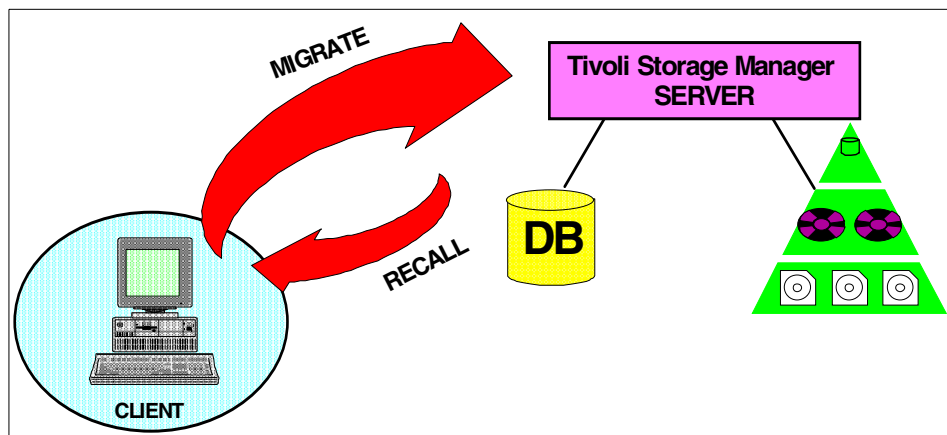
Figure 93. Tivoli Storage Manager Space Manager

See *Using ADSM Hierarchical Storage Management*, SG24-4631, or *ADSM V3R1 UNIX Hierarchical Storage Management*, SH26-4083, for more information about HSM.

Migration and **recall**, discussed in the following sections, are two complementary functions used with Tivoli Space Manager.

12.3.1 Migration

With migration, new files and the files you use frequently remain on your client, and the files you use less often (called *inactive files*) are automatically migrated to the Tivoli Storage Manager server. Thus, HSM frees space on your Tivoli Storage Manager client. You can:

- Specify that files be prioritized for migration according to their size and the number of days since they were last accessed.

- Control client disk occupancy with high and low thresholds. For example, when a file system becomes more than 80% full, begin to migrate inactive files until the occupancy reaches 50%.

- Specify not to migrate a file until it has a backup. Otherwise, if you migrate a file and then take a backup, Tivoli Storage Manager copies the file directly from Tivoli Storage Manager storage and does not require a recall before the backup.

- Manually select files to be migrated. This action is called *selective migration*.

Files migrated to Tivoli Storage Manager storage are always visible to, and easily accessible by, the user and applications. The header of the file, called the *stub file*, remains locally on the client's disk, so that the file is always visible.

The management of migrated files is controlled by policies. Tivoli Storage Manager does not require a separate storage pool for space-managed files, but a separate storage pool is recommended. Clients are likely to require fast access to their space-managed files, and therefore you may want to have those files stored in a separate storage pool that uses your fastest storage.

12.3.2 Recall

When you access a migrated file, HSM automatically recalls it back to your local disk. You may experience a slight delay before accessing the data: A tape mount is usually requested.

You can use several kinds of recall:

- **Transparent recall** — Recall the files from the Tivoli Storage Manager server and store them locally on the client's disk.

- **Migrate-on-close** — Recall a migrated file temporarily. If you do not modify the file, HSM automatically changes it back to a migrated file.

- **Read-without-recall** — Read a migrated file from Tivoli Storage Manager storage without storing it back on your local disk; put it into the memory of the client.

- **Selective recall** — Manually select files to be recalled.

HSM ensures the integrity of your data and provides the same level of security for migrated files as that provided by your operating system files.

12.4 Policy control

Policy control allows different types of data to be treated differently.

You specify, centrally, a set of rules based on user requirements. The rules define how the client files are managed on the Tivoli Storage Manager server.

12.4.1 Storage pools

Each storage pool represents a collection of storage volumes sharing the same media type. For example, a storage pool associated with a device for 3590 tape contains only 3590 tape volumes. You can specify the characteristics of storage pools, such as whether *scratch* volumes are used.

12.4.1.1 Primary storage pool

When a client node backs up, archives, or migrates data, the data is stored in what is called a *primary storage pool*.

Some characteristics of primary storage pools are these:

- They have a hierarchical structure (see Figure 94). Backups, archives, or migrated files can start on disk or go directly to tape. Tivoli Storage Manager controls the movement from one storage pool to the next, according to high and low storage-pool thresholds.

- They can use random access storage (disk device) or sequential access storage (tape, cartridge, optical devices).

- When a user requests to restore, retrieve, or recall, the file is obtained from a primary storage pool, whose volumes are always located onsite.
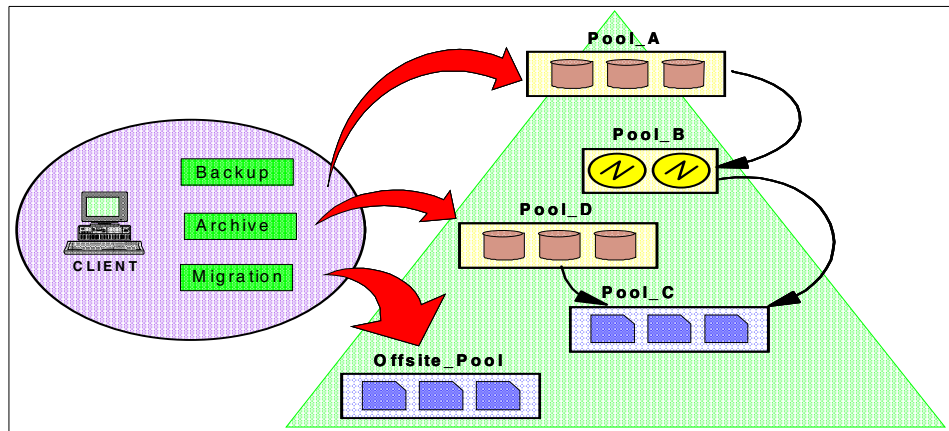
Figure 94. Tivoli Storage Manager primary storage pool

12.4.1.2 Copy storage pools

When an administrator backs up a primary storage pool, the data is stored in a *copy* storage pool.

Some characteristics of copy storage pools are these (see Figure 95):

- One copy storage pool can contain the copies of several primary storage pools.
- A copy storage pool can use only sequential access storage.
- Copy storage pool volumes can be moved offsite and still be tracked by Tivoli Storage Manager. A backup of the Tivoli Storage Manager database should also be stored offsite, as a means of recovering from an onsite disaster.

Note that you need two tape drives if you want to copy from tape to tape.
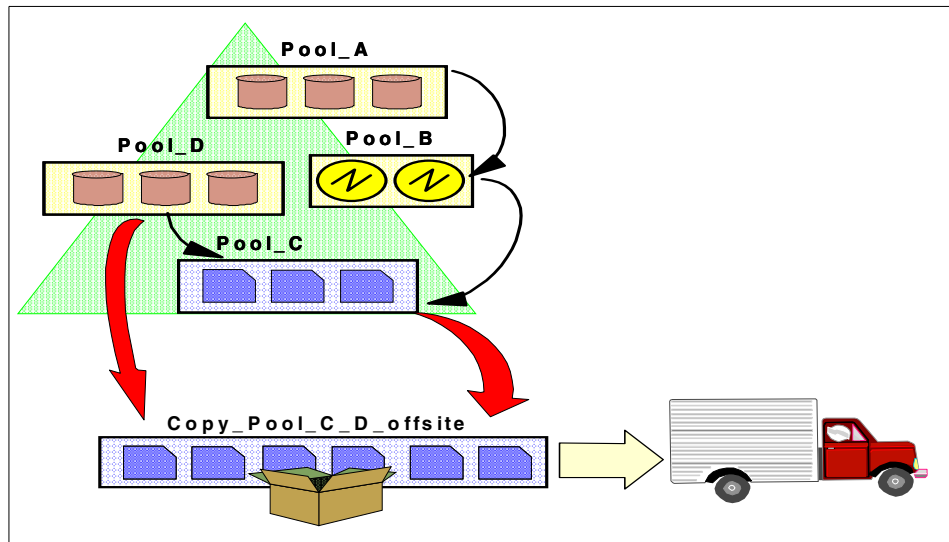
Figure 95. Tivoli Storage Manager copy storage pool

12.4.1.3 Space reclamation

Space reclamation ensures that the tapes are used as efficiently as possible.

Space on a sequential volume becomes reclaimable as files are expired from the volume. Files become obsolete because of aging, or limits on the number of versions of a file, or deletion by an operator.

Figure 96 shows how tapes go from empty, to filling, to 100% full. Over time, data is deleted or expired from the tape, and the tape has unused space. With Tivoli Storage Manager, you can specify a reclamation threshold. When the percentage of expired data on a tape exceeds the reclamation threshold, Tivoli Storage Manager uses the current filling tape to move the remaining data. The now emptied tape can be reused.

Reclamation also works in single drive environments. The remaining data is first copied from the reclaimed tape to the disk, and then migrated from the disk to the current filling tape. Of course, single drive reclamation takes more time than two-drive reclamation.
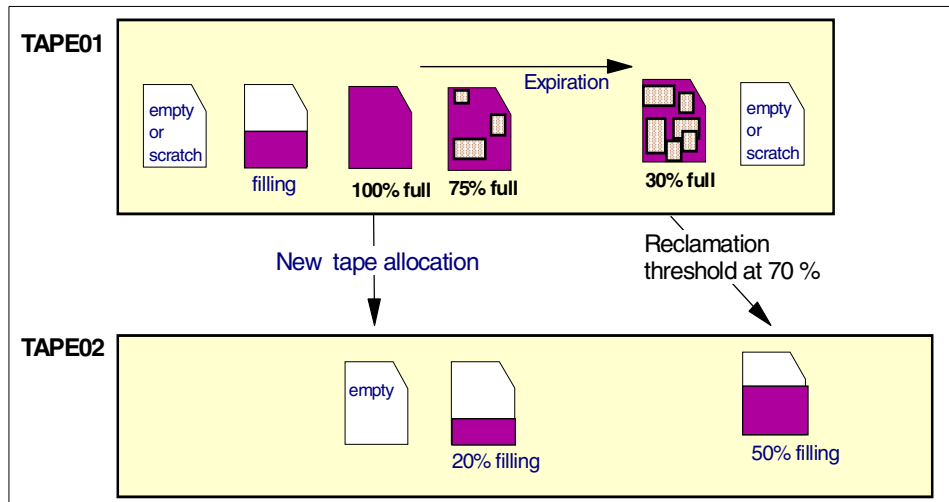
Figure 96. Tivoli Storage Manager reclamation process

12.4.2 Policy domains

With Tivoli Storage Manager, clients are grouped into logical collections, called *policy domains*. A client belongs to one domain, but you can move it to another domain for administrative reasons.

You should define domains based on your own criteria:

- One domain for each major platform; for instance, UNIX, or desktop workstations

- One domain for each major application

- One domain for each geographical or functional site

- One domain for each administrator team

- One domain for each group of client nodes sets that have similar storage management requirements

For each domain, you define specific management classes to treat the client data differently (see Figure 97).

Recall that the more domains you define, the more complex Tivoli Storage Manager administration becomes. The simplest approach is to have one domain and one management class.
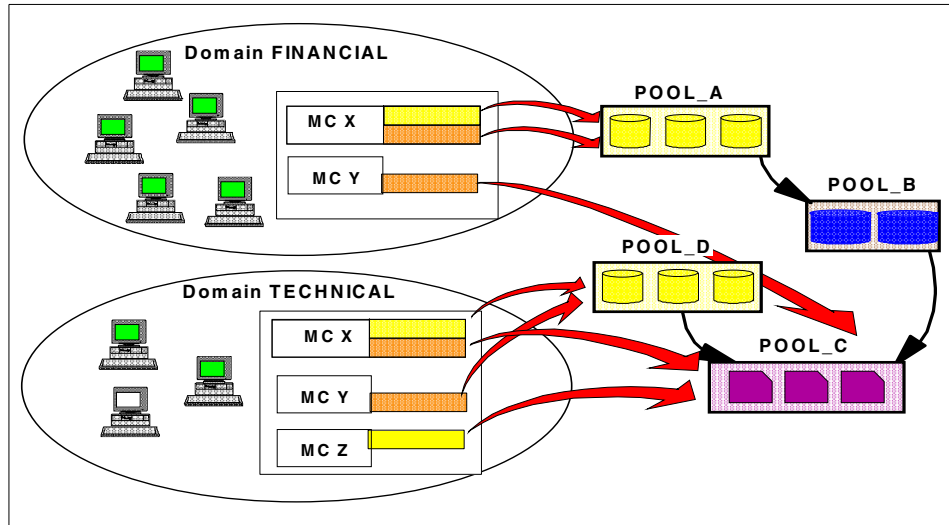
Figure 97. Tivoli Storage Manager domains and management classes

12.4.3 Management classes and client options files

Management classes define how to manage backed up, archived, or migrated data. The parameters for a management class are described in *copy groups*.

As shown in Figure 97, a management class can contain one backup copy group, one archive copy group, both a backup and archive copy group, or no copy groups. The Tivoli Storage Manager parameters are directly defined in the management class rather than in a separate copy group.

12.4.3.1 Backup copy group
The backup copy group controls Tivoli Storage Manager's backup processing of files associated with its management class.

A backup copy group determines:

- The number of backup versions of files the server has to keep

- The number of days the server must keep the older backup versions of files (all versions except the most recent)

- The number of backup versions of files the server must keep, if the files have been deleted from the Tivoli Storage Manager client, and the number of days to keep these remaining copies

- How to handle files currently in use during the backup (back up the file or not)

- The primary storage pool where the server stores backup versions of files

- The number of days before a new backup can be requested for a file. This prevents incremental backups from occurring too often.

12.4.3.2 Archive copy group

The archive copy group controls Tivoli Storage Manager's archive processing of the files associated with its management class. Each archive copy is its own entity; there is no versioning as there is in backup. An archive copy group determines:

- The number of days the server must keep archived copies of files

- How to handle files that are in use during archive (archive the file or not)

- The primary storage pool where the server stores archived files. It can be the same as, or different from, the backup storage pool.

12.4.3.3 Hierarchical Storage Management

The HSM management class parameters define how the client's files are migrated to storage pools. The HSM characteristics are these:

- The primary storage pool where the server stores the migrated files. It can be the same as a pool used for archiving or backing up.

- The number of days the file must remain unused to become eligible for migration because of inactivity.

- Whether or not Tivoli Storage Manager migrates only the files that currently have a backup

- The type of migration:
    - Automatic file migration due to threshold overflow
    - Selective file migration requested by command

12.4.3.4 DSM options file

A Tivoli Storage Manager client can use different management classes, depending on which operation it performs and which data it uses. The client can:

- Back up or migrate files, using the default management class

- Back up or migrate files, using the management class associated with these files in the client option file (*dsm.opt* and *include* option)

- Archive, using the default management class

- Archive, using the management class specified in the command at the time of the archive request

The flexibility of Tivoli Storage Manager enables you to keep the management of your environment as complex or as simple as you want. To keep things simple, you can have one management class be the default and manage all data in the same way.

12.4.4 Tivoli Storage Manager database

The Tivoli Storage Manager database is the most important memory of the Tivoli Storage Manager server. It should be mirrored and backed up daily to tape.

The Tivoli Storage Manager database contains:

- Pointers to the locations of all client files in the Tivoli Storage Manager storage pools. The client data itself is stored in storage pools, not in the database.

- Information about client nodes and administrators.

- Policies and schedules.

- Information about server operations (for example, the activity log).

The server recovery log records the changes made to the database. It is used to restore the database to a consistent state and to maintain consistency across server start-up operations. The recovery log contains information about updates that have not yet been committed to the database.

12.4.5 Centralizing events

Logs allow you and the user to verify how Tivoli Storage Manager works. The Tivoli Storage Manager logs are located on the server and on the client. Centralizing these logs in one place would make it unnecessary to connect to each client to read the logs.

The Tivoli Storage Manager server logs are stored in the Tivoli Storage Manager database. The logs are:

- **Activity log** — Reports the internal activity of the server
- **Event log** — Reports the status of the scheduled events

The Tivoli Storage Manager clients also have logs, stored locally in files on each client. These logs are:

- **Error log** — Reports all the errors that occur on the Tivoli Storage Manager client (*dsmerror.log* file)

• **Schedule log** — Reports all the errors and events that occur in sessions scheduled by the server to the client (*dsmsched.log* file).

To centralize the logs, the clients send events to the server (see Figure 98). The server then filters these events, and eventually stores or forwards them to a central point.
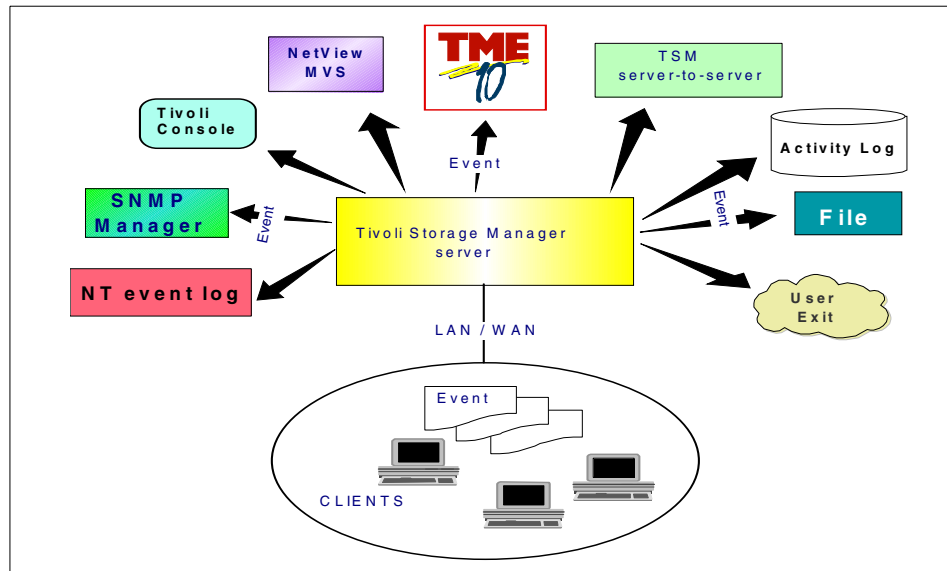
Figure 98. Tivoli Storage Manager client event logging

The central point can be the activity log of the server, the Tivoli Storage Manager console, the Tivoli/Enterprise Console, NetView for MVS, a user exit, a file, the Windows NT event log, a server-to-server event log, an SNMP manager (such as NetView for AIX, CA Unicenter, or HP OpenView).

The concept of redirecting events to external interfaces allows Tivoli Storage Manager's implementations to be integrated with systems management reporting tools.

12.5 User interfaces

User interfaces provide access to Tivoli Storage Manager. Clients usually need to store and recover the data, and the server needs to be administered.

12.5.1 Client interfaces

Client interfaces enable you to manage the Tivoli Storage Manager client.

The Tivoli Storage Manager backup-archive client has a *command-line interface* (CLI) and a GUI (see Figure 99), and a Web-based interface that allows remote graphical management. You can perform more functions from the command line than from the GUI.

Figure 99. Tivoli Storage Manager backup client GUI

IBM also provides a new Web-based backup and an administrative for backup client interface. The Web client consists of new client components that must be installed and configured before the client can be accessed with a Web browser.

The Web client is installed with the backup-archive client package. It is not a separate package as it was in prior versions. The Web client consists of two new processes on the client workstation:

• The client acceptor is an HTTP daemon that serves the Web client Java applet to the Web browser.

- The remote client agent performs the client functions initiated with the Web client interface. The agent does not have to be running all the time. The acceptor daemon starts the agent when client functions are initiated through the Web client interface.

12.5.2 Administrative interfaces

The Tivoli Storage Manager administrative interfaces enable you to manage the Tivoli Storage Manager server.

The administrative client has a CLI and a GUI. The administrative GUI is Web-based (see Figure 100) and enables you to connect to a Tivoli Storage Manager server from any workstation with a Web browser such as Netscape Communicator.
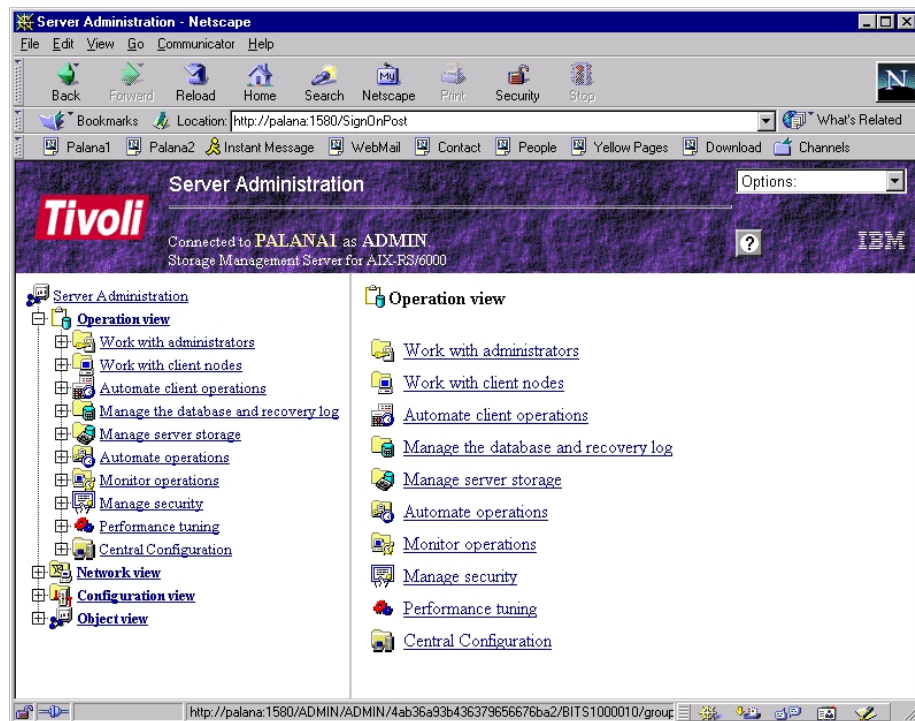
Figure 100. Tivoli Storage Manager Web-based administrative GUI

12.6 Scheduling

You can start Tivoli Storage Manager backup processes through a user request or a central scheduler.

You can start the request on the client, using the backup-archive GUI, or the interactive CLI, or the command line in batches scheduled by a local client scheduler.

You can also start the request in the Tivoli Storage Manager server central scheduler. In this way, you do not have to remember to initiate a backup. At the scheduled start time, the server contacts the client and starts the scheduled action.

12.7 Other functions

Tivoli Storage Manager has some additional functions that provide access to client database backups, communication between Tivoli Storage Manager servers, and support for a wide variety of server and client hardware and software.

12.7.1 Backup/Archive client multi-threading

Until now the Tivoli Storage Manager Backup/Archive Client does not exploit the multi-threading capabilities of modern operating systems. The new Tivoli Storage Manager client starts multiple threads, in order to transfer data more quickly. This is achieved by starting multiple session from the client to the server. This functionality will work for the following client functions:

- Backup client function
- Archive client function

This new function is completely transparent to the user. There is no need to switch it on or off. The Tivoli Storage Manager client will decide if a performance improvement can be gained by starting an additional session to the server.

The client will determine how many sessions will be started to transfer the data according to the speed of the current workload.

The advantages of this implementation are good load balancing capabilities (can be used on a single file specification) and estimated good performance gains on backups that send large amounts of data to the server.

There will always be two sessions, even if multithreading is not really used. That is because there will always be at least one control and one data thread/session.

12.7.2 Tivoli data protection for database backups

Many techniques are used to back up data stored in relational database management systems (RDBMSs); for example, disk mirroring, database export, offline or online backup, and full or partial database backup.

With Tivoli Storage Manager, you can back up any database or application that stores its data in files, if the files are closed.

With Tivoli Storage Manager you can back up some databases online, without stopping them. A Tivoli Storage Manager connection, known as *Tivoli Data Protection,* makes the link between the Tivoli Storage Manager server and your database (see Figure 101).
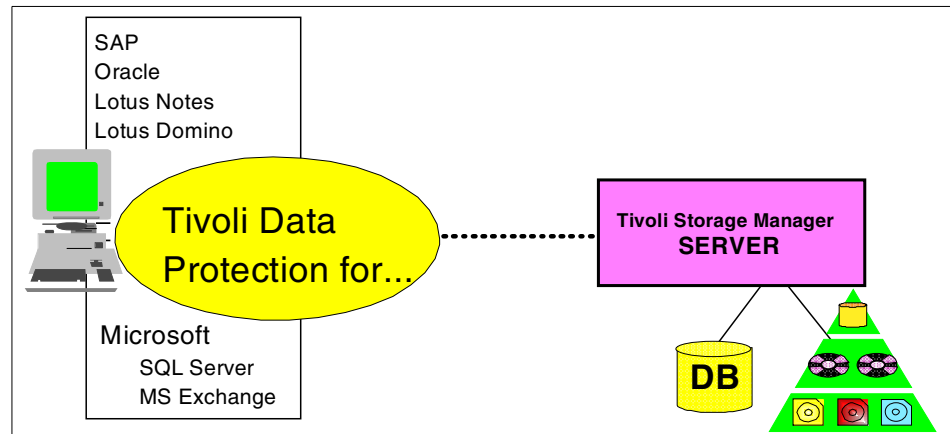
Figure 101. Tivoli data protection software

For more details, see *Using ADSM to Back Up Databases*, SG24-4335 and *Using ADSM to Back Up Lotus Notes*, SG24-4534.

12.7.3 Clustering Tivoli Storage Manager servers

Tivoli Storage Manager server-to-server communication enables you to store backup, archive, and HSM-managed data with multiple Tivoli Storage Manager servers, using the network.

Tivoli Storage Manager server-to-server communication (see Figure 102) enables you to:

- Store the client's data on a Tivoli Storage Manager server other than the default server.

- Back up the Tivoli Storage Manager database to another Tivoli Storage Manager server.
- Back up primary storage pool data to a copy storage pool on another Tivoli Storage Manager server.
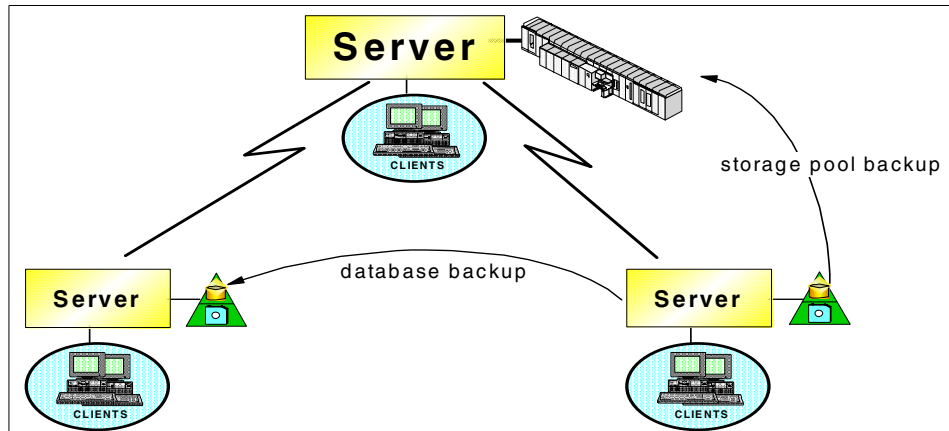
Figure 102. Tivoli Storage Manager server-to-server communication

The location of the data is transparent to the client. You can define policies that direct the data to different locations for security, recovery, or capacity reasons. In conjunction with the Tivoli Storage Manager Disaster Recovery Manager (DRM) feature, the Tivoli Storage Manager server-to-server communication feature provides disaster protection for the Tivoli Storage Manager database and storage pools. The tapes do not have to be manually moved to and from an offsite location.

See *ADSM Server-to-Server Implementation and Operation*, SG24-5244 for details.

12.7.4 Tivoli Storage Manager tape sharing

Tivoli Storage Manager introduces the sharing of tape libraries between servers to address the following issues:

- Use tape resources more efficient, until now it was only possible to share certain drives between two hosts.
- Optimizing the use of scratch tapes by sharing them across multiple Tivoli Storage Manager servers.

The basic architecture that is used for this support is based on the External Library Manager (ELM) interface that is currently implemented in the prior

ADSM server. This architecture differentiates between the tape drives and the robotic device (the library). The drives are used by any hosts, but only one controls the robotic device.

The functionality of sharing tape libraries is split into two components: the library manager and the library client.

- The task of the library manager is to physically control the library. All the communication with the library is done by this server. The library server can also use the library for its own purposes, without being a library client at the same time (no library client is necessary on the manager).

- The library client uses the library server for the handling of the physical hardware. The client send the requests to the library manager, and afterwards reads or writes to the tape drive he was given access to.

12.7.5 Multiplatform support

Because Tivoli Storage Manager manages the storage in a heterogeneous environment, it supports many server platforms, client platforms with or without HSM, and database backups.

For more details and up-to-date information, see the Web site:

`http://www.storage.ibm.com/software/adsm/adsercli.htm`

The platforms supported for different Tivoli Storage Manager roles are as follows.

- **Tivoli Storage Manager server platforms**

    - Microsoft Windows NT Version 3.51 and 4.0
    - IBM AIX 3.2.5, 4.1.4, 4.1.5, and 4.2.1
    - HP-UX 10.10, 10.20, and 11.0
    - Sun Solaris 2.5.0, 2.5.1, and 2.6
    - OS/400 V3 R1, V3R2, V3R6, V3R7, V4R1, and V4R2
    - OS/390 R1, MVS/ESA SP V5 R1, MVS/ESA SP V4, MVS V3 R1.3
    - VM/ESA Version 1.1.5 370 Feature, 1.2.2, and 2.1.0
    - VSE/ESA Version 1.3

- **Tivoli Storage Manager-supported storage devices**

    - 8mm and 4mm drives and autochangers
    - QIC tape drives and libraries
    - SONY DTF tape drives
    - DLT tape drives and libraries
    - SCSI attached tape transports and libraries, and Library Dataservers
    - Parallel attached tape transports, and Library Dataservers

- ESCON attached tape transports and Library Dataservers
- Optical drives
- Optical Jukeboxes (Native SCSI Rewritable, and WORM)
- Tivoli Storage Manager/STK ACSLs supported drives and libraries

- **Tivoli Storage Manager client platforms**

  - Macintosh Operating System 6.0.7 and later
  - Auspex (via the SunOS client)
  - Bull DPX/2 300 B.O.S. Release 2.00
  - Bull DPX/20 (via the AIX client)
  - CRAY UNICOS (via the AIX client)
  - Data General DG/UX
  - Digital OpenVMS (via SSSI's ABC OpenVMS Client for Tivoli Storage Manager)
  - Digital ULTRIX 4.3, 4.4, and 4.5
  - Digital UNIX Version 3.2 (DEC OSF/1) and 4.0
  - DOS 5.0 and later
  - Fujitsu UXP/V and UXP/M
  - HP-UX 8.07, 9.01, 9.03, 9.04, 10.01, 10.20, 11.0
  - AIX 3.2, 4.1.4, 4.1.5, 4.2.1 and 4.3
  - IBM OpenEdition MVS/ESA SP Version 5.1 and OS/390 SMP/E Rel.8.0
  - IBM OS/390 UNIX System
  - IBM OS/2 Warp and Lan Server with 386 HPFS
  - Linux (this client is offered on an unsupported basis only)
  - Microsoft Windows 3.1
  - Windows 95 & 98 & 2000
  - Microsoft Windows NT Version 3.5.1 and 4.0, on Digital AlphaServer
  - Microsoft Windows NT Version 3.5, 3.5.1, and 4.0, on Intel-based
  - NCR UNIX SVR4 MP-RAS Release 2.02, 2.03, 3.0, 3.01 and 3.02,
  - Novell NetWare 3.11, 3.12, 4.01, 4.02, 4.1, and 4.11
  - NEC EWS-UX/V 4.2, revision 9.1, 9.2, 10.1, 11.x, 12, 13
  - SCO UNIX 386 R 3.2 V 4.2 and 5.0, Open Desktop 2.0, 3.0, and OpenServer 5
  - Sequent PTX UNIX Release 2.1.6, 2.1.7, 4.1.3, 4.2.3, 4.4.1
  - Siemens Nixdorf SINIX-Z 5.4.1, SINIX-N, SINIX-P 5.4.2, SINIX UNIX 5.4.3
  - Silicon Graphics IRIX UNIX 5.2, 5.3, and 6.2
  - Sun Solaris 2.3, 2.4, 2.5.0, 2.5.1, and 2.6
  - Sun SunOS 4.1.3
  - Tandem Guardian (NSK)

- **Tivoli Storage Manager HSM client platforms**

  - AIX 3.2, 4.1, 4.2.1 and 4.3

- Silicon Graphics IRIX UNIX 5.2, 5.3, and 6.2
- Sun Solaris 2.5.0 with UFS, and 2.5.1 with VFS using DMAPI
- Silicon Graphics IRIX UNIX R6.2 with XFS using DMAPI
- Windows NT provided by OPEN/stor (reseller)
- Netware provided with AvailHSM from Wang Software (reseller)

- **Tivoli Storage Manager applications to back up databases**

  - Tivoli Data Protection for Lotus Notes on AIX, Windows NT, OS/2
  - Tivoli Data Protection for Microsoft Exchange Server
  - Tivoli Data Protection for Microsoft SQL Server
  - Tivoli Data Protection for Oracle Backup on AIX, HP-UX, Sun Solaris, Windows NT
  - Tivoli Data Protection for Lotus Domino - new product providing support for the new Lotus Domino 5 APIs.
  - Tivoli Data Protection for Informix.
  - Tivoli Data Protection for SQL-BackTrack for Oracle and Sybase
  - Tivoli Data Protection for SAP/R3 with Backint for Oracle, and ADINT for ADABAS-D databases

12.8 Specifications

In this section we present specifications to help you try-and-buy or implement Tivoli Storage Manager.

12.8.1 Prerequisites

Tivoli Storage Manager manages most major platforms and works in most environments.

Therefore, before implementing Tivoli Storage Manager, you have to consider the operating systems and hardware of your clients and servers, your database versions and platforms, communication protocols, and storage devices.

For all details and up-to-date information, see:

```
http://www.storage.ibm.com/software/adsm/adsercli.htm.
```

12.8.2 Priced features

Tivoli Storage Manager's features are priced separately (see Table 8 on page 251). You must register a new license in Tivoli Storage Manager if you want to add support for any of the following features:

- CLIENTS: Additional backup-archive clients

- NETWORK: To allow Tivoli Storage Manager server to use the communication protocols (stand-alone implementation if not)
- SPACEMGMT: HSM clients
- DRM: Disaster Recovery Manager
- ADVANCED DEVICES: Extended storage device support
- OPENSYSTEMS: OpenSystem environment support

In the example described in Table 8, the two servers are communicating:

Table 8. Tivoli Storage Manager: MVS and Windows NT server examples

Environment to manage	Tivoli Storage Manager features
MVS server with a 2-rack IBM 3494 library	Tivoli Storage Manager for MVS Base Server NETWORK SPACEMGMT ADVANCED DEVICES
Windows NT server with an IBM 3570 library	Tivoli Storage Manager for Windows NT Base Server NETWORK
18 Windows NT clients	18 CLIENTS
10 Netware clients	10 CLIENTS
6 AIX clients with HSM managed by the MVS server	6 CLIENTS
2 Sun Solaris clients	2 CLIENTS
3 HP-UX clients	3 CLIENTS
5 Lotus Notes on Windows NT	5 Tivoli Data Protection for Lotus Notes on Windows NT
3 Oracle clients on AIX	3 Tivoli Data Protection for Oracle on AIX
3 AIX SAP/R3 databases with Oracle	3 CLIENTS 3 BACKINT for Oracle

Chapter 13. Data Facility Storage Management Subsystem

This chapter explains the Data Facility Storage Management Subsystem/MVS (DFSMS).

Reading through these topics, you will discover many functions in the DFSMS product that are similar to those in the Tivoli Storage Manager, as explained in Chapter 12, "Tivoli Storage Manager" on page 223. This could be a good opportunity to consider consolidating your storage management teams from both worlds, open systems and S/390, for a unique and more efficient storage management team. See 16.1, "Understanding enterprise storage administration" on page 305 for details.

Although DFSMS provides various functions, including access methods, file systems, catalog management, and utilities, as a part of the OS/390 operating system, we will focus on storage management functions in this chapter.

To learn about the latest improvements in existing and new DFSMS functions, see the *DFSMS/MVS 1.5 Presentation Guide*, SG24-5356. Here is a summary of features in this new release:

- Improved performance
    - Hierarchical File System access
    - Shared catalog in a sysplex
    - DFSMShsm volume backup
    - DFSMSdss DEFRAG RVA enhancement
- Enhanced sysplex exploitation
    - Multiple DFSMShsm HSMplexes
    - Secondary host promotion when the DFSMShsm operation on the primary host fails
    - Access to OAM objects from any system in the sysplex
    - Integrity of HFS mounts in a shared environment
- Enhanced DFSMSrmm removable media management
    - Provides an API for read/write access
    - Supports separate retention and vaulting policies
    - Supports environments with unshared catalogs
    - Enables special handling instructions

- Improved storage administration functions

  - Provide additional function in NaviQuest batch

  - Support master catalog alias resolution by system

  - Support DFSMSdss logical operations at storage group level

  - Support DFSMShsm volume backup operations at the storage group level

- Enhanced application growth enablement

  - Supports multivolume HFS

  - Allows a large number of open data sets for IMS and CICS TS Subsystems

  - Allows HSM Control Data Sets to be larger than 4 GB per cluster

- Expanded data management capabilities

  - Extended Format support for all remaining VSAM data set organizations (ESDS, RRDS, LDS, VRRDS), including support for:

    - Extended Addressability (EA)
    - Partial release
    - Candidate volume space amount
    - System Managed Buffering

  - EA for VSAM LDS that supports DB2 table spaces larger than 1 TB — available in DB2 Version 6

- Improved interoperability

  - Supports ISO/ANSI Version 4 tapes

You can also visit the Web site:

```
http://www.storage.ibm.com/storage/software/sms/smshome.htm
```

13.1 Overview

System-managed storage (SMS) is a set of storage management functions that DFSMS provides for storage resources such as data sets, DASD, tapes, and optical devices attached to S/390 platforms. SMS creates a policy-driven, highly automated storage management environment for using storage resource more efficiently. SMS uses many different components of DFSMS to provide these services.

SMS allows the storage administrator to affect data set attributes, placement, migration and backup by assigning classes and storage groups to new data

sets. Classes and groups are assigned through installation written automatic class selection (ACS) routines you code to ensure your storage management policies.

13.1.1 Key features

DFSMS provides the following key features:

- **Availability management** — Provides functions for taking incremental backups, volume dumps, or application level backups for disaster recovery.

- **Space management** — Provides functions for HSM and other techniques to keep free space.

- **Policy control** — Ensures your storage management policies throughout the SMS complex.

- **User interface** — Provides a common interface called *interactive storage management facility* (ISMF) to storage administrators and users to manage storage resources easier.

- **Scheduling** — Automates availability and space management functions described above.

- **Removable media management** — DFSMSrmm maintains an inventory of the location and contents of tape volumes and controls access to and movement within the tape inventory.

13.1.2 Product components

DFSMS is composed of the following four major components:

- **DFSMSdfp** — Provides basic data access services, including access methods, file systems and device support, as well as providing policy management for data through the SMS subsystem.

- **DFSMSdss** — Provides a mechanism for fast data movement. It makes copies, dumps/restores, converts or prints data sets or volumes. Each function can be invoked through JCL, ISMF panels, API, or DFSMShsm.

- **DFSMShsm** — Provides availability and space management functions for data sets and volumes. It also provides scheduling function to automate these tasks. DFSMSdss is mainly used as a data mover in most of DFSMShsm processes. DFSMShsm provides the additional benefits of automated scheduling and inventory management.

- **DFSMSrmm** — Allows tape volumes and data sets to be system-managed. It provides validations and policy management functions.

DFSMS Optimizer is discussed in 13.8.3, "DFSMS Optimizer" on page 280.

13.1.3 Operating environment

DFSMS runs as a part of the OS/390 operating system on S/390 processors.

13.2 Availability management

DFSMShsm provides a group of functions called *availability management*. The purpose of availability management is to ensure that lost or damaged data sets can be retrieved from a backup taken at a known time. Availability management provides the following functions:

- 13.2.1, "Backup, dump, and recover" on page 257
- 13.2.2, "Disaster recovery" on page 258

Figure 103 shows the different forms of backup copies that DFSMShsm can provide. You can produce each type of backup copy by several means. For example, the local and offsite backup copies can be produced by a combination of incremental backups and full volume backups where the tapes for both have been duplicated. Similarly disaster recovery backup copies can be produced by Aggregate Backup and Recovery Support (ABARS).
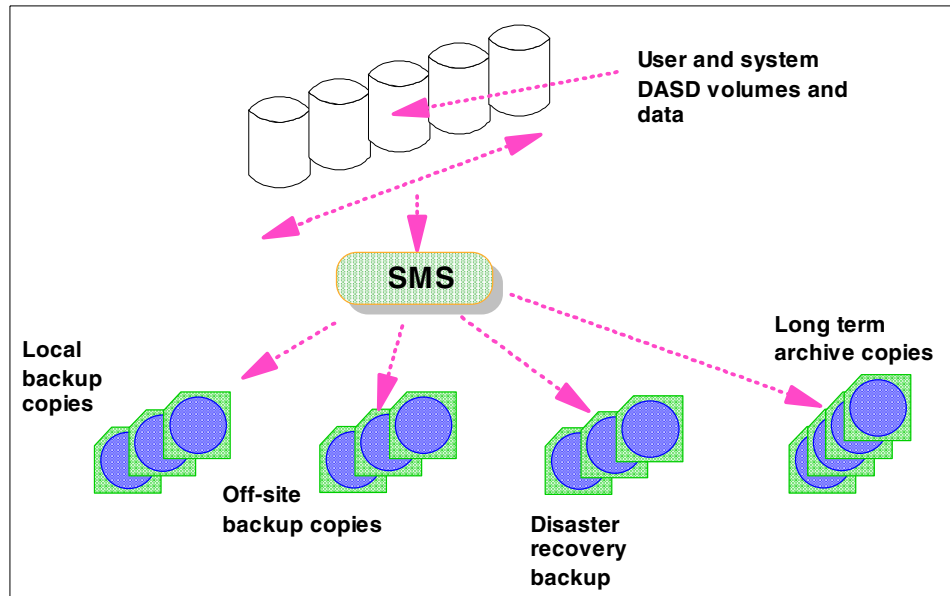
Figure 103. DFSMShsm backup processing

13.2.1 Backup, dump, and recover

This section covers the backup, dump, and recover processes.

13.2.1.1 Backup

DFSMShsm backup processes copy data from an SMS-managed or DFSMShsm-managed volume to a DFSMShsm-owned daily backup volume. This copy is called a *backup version*. DFSMShsm creates and manages up to 100 backup versions per data set and supports incremental backup. After DFSMShsm successfully backs up a data set, it resets the changed indicator in Volume Table of Contents (VTOC) entry for that data set. DFSMShsm minimizes the space occupied by the data on the DFSMShsm owned backup volumes by compacting it.

You can specify how frequently to back up data sets on a data set basis by assigning an appropriate management class.

The BACKVOL command can be specified with a new STORGRP parameter that allows all volumes in the named storage group(s) to be processed for incremental backup or full volume dump operations.

13.2.1.2 Dump

DFSMShsm dump produces a DFSMSdss physical dump of an entire volume. This is called a *dump copy* and is stored on a dump volume. The difference between dump and backup is that the dump function backs up the entire allocated track images on a volume, whereas the DFSMShsm backup function backs up individual data sets based on record basis. Dump copy reduces the time required to perform volume recovery.

The number of dump copies and their retention is determined by the dumpclass(es) assigned to a volume or storage group. You can dump different groups of volumes on different days with different periods for the number of days between dumps. You can also retain different dump copies for different periods by using different dump class retention periods.

13.2.1.3 Recover

Recover processing can be performed by using incremental backups or dump copies as input or you can use a combination of both methods.

Data set recovery selectively restores data from specific backup copies. Generally, DFSMShsm will recover data from the most recent dump copy or backup version available. You can also rename the data set during recovery without impacting a currently existing data set.

Volume recovery is needed when an entire DASD pack has been lost or damaged, or when a significant amount of the data has become inaccessible. Recovering a single volume will impact data sets that have extents on this and other volumes, as well as applications that have data sets on this and other volumes.

To recover volumes, three methods can be used:

- Full volume restore from dump copies
- Volume recovery from incremental backup versions
- Full volume restore with update from incremental backup copies

13.2.2 Disaster recovery

Aggregate backup and recovery support (ABARS) is another DFSMShsm backup and recovery function for disaster recovery. It operates independently of other availability management functions. ABARS processing is centered around an aggregate. An aggregate is defined as a group of data sets and their associated control information.

Aggregate groups are defined using ISMF option 9. An aggregate includes:

- **Selection data set** — This data set contains the data set filters that determine what data is included in the aggregate. Data sets can be both included and excluded using the selection lists. Data sets can be selected for processing, selected to be allocated at the recovery site, or data on moveable media can be selected to physically accompany the aggregate.

- **Instruction data set** — This data set contains a list of instructions that describe what to do with the aggregate as it is recovered.

- **Output data set** — This data set will be used as for ABARS processing.

- **Management class** — This class will be associated with the aggregate and determines when the aggregate expires.

- **Count** — This value determines the number of concurrent copies produced when a new aggregate generation is produced.

Aggregate processing is accomplished by two DFSMShsm commands: ABACKUP, which backs up an aggregate group; and ARECOVER, which recovers data sets that were previously backed up using the ABACKUP command. ARECOVER processing can also be used to selectively recover data sets from an aggregate. For ABACKUP processing the SMS subsystem must be active on the system processing the ABACKUP command. SMS is not required, but is strongly recommended to be active on the system running an ARECOVER.

The output of a successful aggregate backup command is:

- **Data files** — These files contain copies of the data sets that have been backed up. Two separate files can be produced: one for data sets from primary volumes, and a second for migrated data.

- **Control file** — This file contains control information needed by aggregate recovery to verify or to recover the application's data sets. The control file also specifies which ACCOMPANY and ALLOCATE data sets should be cataloged or allocated at the recovery site.

- **Instruction/activity log file** — An optional file that contains the instruction data set. The instruction data set contains the instructions or information that may be needed at the recovery site. A copy of the activity log is also written to this file if the activity log is directed to DASD.

To perform aggregate recovery, you must have the control file, data files, and instruction/activity log file created by a previous aggregate backup. When you issue the ARECOVER command, DFSMShsm creates:

- The recovered data sets — Data sets recovered from the aggregate backup tape data file.

- An activity log — A log containing messages written during aggregate recovery processing.

- A DFSMShsm statistics record — A record containing the statistics for aggregate recovery processing.

- Updates to the DFSMShsm ABARS control records in the HSM CDSs, showing the CPU time for processing ARECOVER requests.

Aggregate recovery can also create the following:

- **Conflict resolution data set** — This data set is used as input to a subsequent ARECOVER command. The conflict resolution data set contains names of data sets that the ARECOVER verification process determines to have like-named conflicts that were not resolved by the RECOVERNEWNAME parameters, the DATASETCONFLICT parameter, or the ARCCREXT installation exit. For each such data set encountered, an entry is placed into the conflict resolution data set, with a default action of BYPASS.

- **Restart data set** — This data set is created when the recovery of data sets does not complete successfully. The restart data set contains the names of all the data sets recovered successfully. When an aggregate recovery is reissued, the restart data set ensures that data sets already processed are not recovered a second time. If you want to start aggregate recovery again from the beginning, delete the restart data set.

Unlike Tivoli Storage Management, DFSMS does not provide a dedicated archive process. Data archiving can be achieved using ABARS. DFSMShsm manages the inventory of individual versions of archives taken by ABARS, based on the management class assigned to an aggregate.

See *DFSMShsm ABARS and Mainstar Solutions*, SG24-5089 for more details.

13.3 Space management

Space management allows you to keep primary DASD space available for users in order to meet the service-level objectives for your system. The purpose of space management is to manage your primary DASD storage efficiently.

The following components of DFSMS provide space management functions:

- DFSMShsm provides the following space management functions based on management class attributes:

    - Deleting data sets that have expired.

    - Releasing unused allocated space from data sets.

    - Migrating data sets from SMS-managed or level 0 volumes to lower storage hierarchy.

    - Recalling migrated data sets to the level 0 devices when needed.

13.3.1 Migration and recall

DFSMShsm migrates inactive data sets to lower storage hierarchies such as migration level 1 or migration level 2 volumes. DFSMShsm does the following:

- Expires data sets that have passed their expiration dates.

- Reduces the space occupied by data by freeing any unused portion.

- Migrates unused data sets to locations where they can be stored less expensively, or in a form that saves space on the volume.

DFSMShsm performs space management according to policies defined and administered by the storage administrators. These policies are applied differently to SMS and non-SMS managed data. All data sets on non-SMS volumes are treated equally by DFSMShsm space management but data sets on SMS volumes are managed according to their management class.

DFSMShsm automatically recalls needed migrated data sets to level 0 volumes.

The latest release, DFSMS R10, provides fast subsequent migration. The idea here is that data sets that are recalled from ML2 tape, but not changed or deleted, can be reconnected back to the original ML2 tape. This eliminates the unnecessary data movement resulting from re-migration, and may reduce the need to perform RECYCLE processing against these tapes.

This process is illustrated in Figure 104.
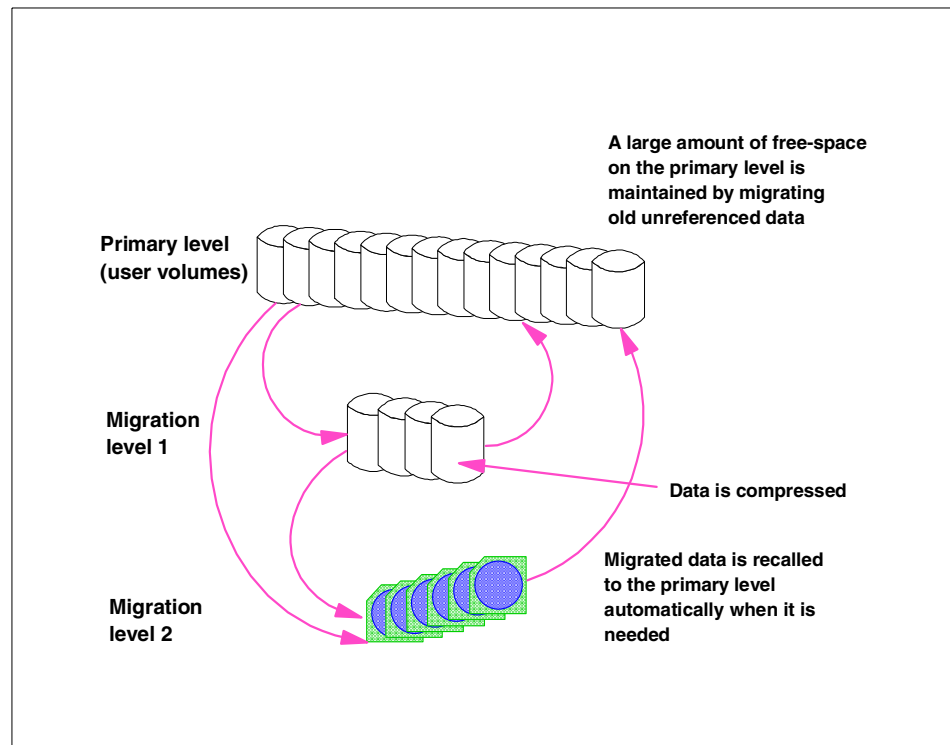
Figure 104. Tasks performed by HSM space management

13.3.2 Other space management functions

Other space management functions that DFSMS provides are the following:

- **Expiration** — Deletes expired data sets.
- **Partial release** — Reclaims allocated but unused space in data sets.

- **Compressing partitioned data sets** — Retrieves unusable space from partitioned data sets (PDS).

- **Defragmentation** — Consolidates the free space extents on volumes.

- **Allocation control** — DFSMSdfp selects target volumes for new data sets to be allocated to balance the usage of volumes.

- **Compressing sequential data sets** — DFSMSdfp compresses active data sets by providing standard or tailored dictionaries.

13.4 Policy control

One of the major tasks of storage administrators is to implement and enforce storage management policies and standards. Having defined standards provides no benefit if there is no way to enforce standards automatically. DFSMS supplies mechanisms for policy control in several components.

13.4.1 Storage group and DFSMShsm owned volume

In an SMS-managed environment, there are two kinds of storage pools implemented for active and inactive data.

Figure 105 summarizes storage pools deployed in DFSMS.
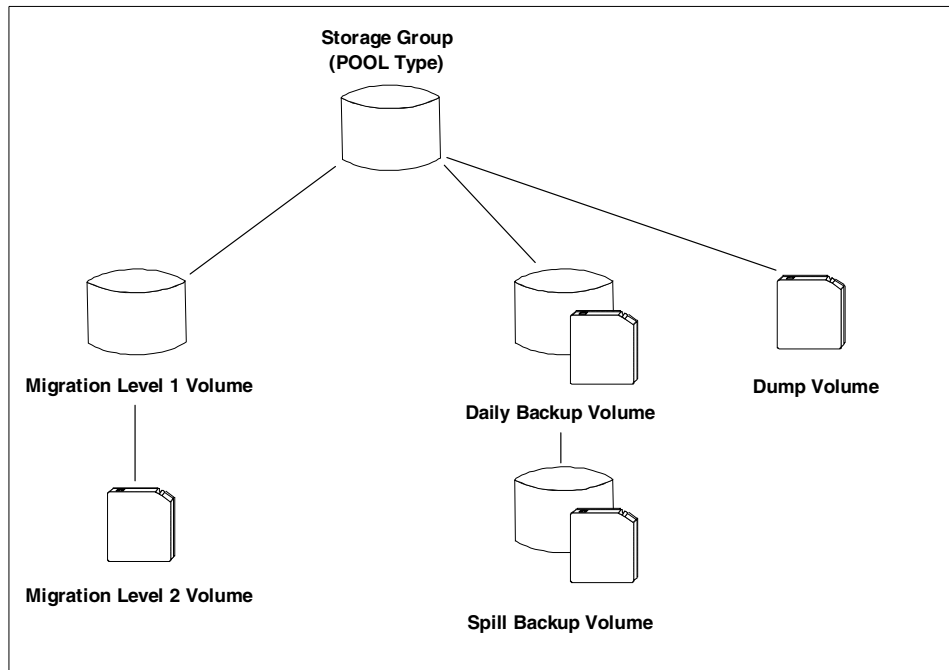
Figure 105. DFSMS storage pools and their hierarchy

13.4.1.1 Storage group

Storage groups represent the physical storage managed by SMS. This storage can be collections of user DASD volumes, which is called POOL type storage group, volumes in tape libraries, optical devices, or virtual input/output (VIO) storage.

Within storage groups you define the thresholds that invoke space management and the thresholds used to spread allocations within a storage group. Different storage groups can be managed with different thresholds. You can also specify which systems are to perform DFSMShsm automated processing against a storage group.

The optimum number of pool storage groups is probably one. However, it is unlikely that an enterprise will be able to reduce the number of pool storage groups to this level, due to factors including differing device capabilities, reliance on physical pool backups, and site specific historical and political reasons.

Storage groups are defined using ISMF option 6.

13.4.1.2 DFSMShsm owned volume

DFSMShsm owned volume is a storage pool for inactive or backup data and is similar to ADSM storage pools. DFSMShsm can have up to two storage hierarchies for migration and backup.

- **Migration level 1 volume**

  Migration level 1 (ML1) volume is a DASD storage pool for migration. You can have only one ML1 pool for each DFSMShsm complex.

- **Migration level 2 volume**

  Migration level 2 (ML2) volume is a storage pool for level 2 migration. DASD or tape device can be used. You can have only one ML2 pool for each DFSMShsm complex.

- **Daily backup volume**

  Daily backup volume is a storage pool for backup. DASD and tape device can be used. You have one daily backup volume pool per each backup cycle.

- **Spill backup volume**

  Spill backup volume is a storage pool for level 2 backup. DASD and tape device can be used. You have only one daily backup volume pool for each DFSMShsm complex.

- **Dump volume**

  Dump volume is a tape storage pool for dump. You have one dump volume pool per each dump cycle.

13.4.2 Management class

Management classes define how your data sets will be treated by DFSMShsm, including whether the data set will be expired at some time in the future, and how frequently it is eligible to be backed up or migrated. For an object, the management class defines the "class transition" attributes of the object as well as expiration and backup criteria. For an aggregate, the management class defines retention and serialization attributes.

Some of the characteristics determined by management class include:

- Incremental backup, how frequently it occurs, and how many copies are maintained.

- Which method is used during incremental backup: regular, concurrent copy, or SnapShot.

- Whether data is eligible to be migrated, and how long it resides at each migration level.
- When data, and its backups, expire.
- Requirements for releasing over-allocated space.
- Generation data group (GDG) information
- Object class transition criteria.

Management classes are defined using ISMF option 3.

The user cannot override any of the individual attributes of the management class, but can request a specific management class to be assigned. The storage administrator can determine whether or not to honor a request for a specific management class.

13.4.3 Data class and storage class

Other classes used by SMS to classify service level for each data set are data class and storage class.

13.4.3.1 Data class
Data class is a collection of physical data descriptors that can be used to determine what a new data set will look like. In general, any value specified in a data class can be explicitly overridden by the allocator of the data set.

Some of the attributes assigned using data class include:

- The record length, format, and organization of the data.
- How much space should be allocated for the data set, both primary and secondary allocation. You specify it in the number of records.
- Over how many volumes this data should be spread, and whether this space allocation can be retried if it is initially unsuccessful.
- Whether the data should be striped.
- Whether this data should be compacted or not; this applies to both tape and DASD data sets.

Having standards for data characteristics determined by data classes provides the following benefits to storage administrators:

- Users need not be aware of data attributes; coding a data class should be sufficient.

- Changes to standards can be implemented without impacting users. You can change the class without needing to change users allocation control statements.

- New technologies can be implemented seamlessly.

End users can override most of the values specified in a data class to tailor the definition more closely to their data's requirements. The storage administrator can also choose to honor a user's explicit specification of a data class on their allocation.

Data classes are defined using ISMF option 4.

13.4.3.2 Storage class

A storage class is a list of storage objectives and requirements. Each storage class represents a list of services that are available to data sets. A storage class does not represent any physical storage, but rather provides the criteria that SMS uses in determining an appropriate location to place a data set or object. SMS uses storage classes to separate data set performance objectives and availability from physical storage. Storage classes are used to determine the following:

- What device type will be selected for a data set, 3380, 3390-3 or 3390-9.

- Whether the device should be allocated to a specific volume or range of volumes.

- Whether the data set is eligible for VSAM record level sharing (RLS).

- What response time is required for the data (whether the data should be cached: always, sometimes, or never.)

- Whether the data set should be allocated on a RAID or mirrored device.

- Whether the data set should be placed on a device that is capable of concurrent copy or SnapShot.

- Whether the data has particular performance requirements that can only be satisfied by data striping or compression.

- Whether access to the data is biased towards read or write or towards sequential or direct processing.

- How buffering should be allocated for VSAM applications using system managed buffering.

If a DASD resident data set is not assigned a storage class, then the data is not SMS managed. Data sets directed to SMS managed tape volumes may

be assigned a storage class; however, they are not SMS managed. SMS only manages tape volumes, not the data sets that are allocated on them.

End users cannot override any of the specific attributes specified within a storage class (unlike data class attribute overrides, which are allowable). As a user, you can specify the name of a storage class on their allocation request. Storage administrator can decide whether you wish the specification of a named storage class to be honored.

Storage classes are created using ISMF option 5.

13.4.4 ACS routines

Automatic class selection (ACS) routines are used to determine the SMS classes and storage groups for data in your SMS complex. There is an ACS routine for each of the three classes and for the storage group. For storage administrators, ACS routines automate and centralize the process of determining SMS classes and storage groups.

ACS routines are written using the ACS programming language, a high-level programming language. The language follows a logical, procedural flow of implementation that consists mainly of filtering criteria, IF/THEN statements, or SELECT/WHEN statements. Using these relational statements, ACS routines determine SMS classes and storage groups according to allocation parameters, data set sizes, object or data set names, and other variables.

Generally, the order of execution for ACS routines is as follows:
1. Data class
2. Storage class
3. Management class
4. Storage group

The management class and storage group ACS routines are executed only if a storage class is assigned to the allocation.

All allocations directed to units that are neither tape nor DASD should be excluded from the SMS-managed environment.

A sample storage class ACS routine is shown in Figure 106.

```
PROC STORCLAS

/***********************************************************************
/
/*              Start of DB2-related FILTLISTs                        */
/***********************************************************************
/
FILTLIST DB2                    INCLUDE(%DB*.**)
FILTLIST DB2_PROD_DIRECTRY      INCLUDE(PDB*.*.DSNDB01.**)
FILTLIST DB2_PROD_CATALOG       INCLUDE(PDB*.*.DSNDB06.**)
FILTLIST DB2_PROD_LOG           INCLUDE(PDB*.**.LOG*)
FILTLIST DB2_PROD_BSDS          INCLUDE(PDB*.**.BSDS*)
/***********************************************************************
/
/*              End of DB2-related FILTLISTs                          */
/***********************************************************************
/
/***********************************************************************
/
/*    Start of DB2 Select                                            */
/***********************************************************************
/
  WHEN (&DSN = &DB2)                        /*  Select DB2 data sets  */
    DO
      SELECT
        WHEN (&DSN = &DB2_PROD_LOG)       /*  Use fast write for    */
          DO                              /*  active logs           */
            SET &STORCLAS = 'FASTWRIT'
            EXIT
          END
        WHEN (&DSN = &DB2_PROD_CATALOG OR  /*  Dual copy for catalog */
              &DSN = &DB2_PROD_DIRECTRY OR /*  directory and boot    */
              &DSN = &DB2_PROD_BSDS)       /*  strap data set        */
          DO
            SET &STORCLAS = 'DBCRIT'
            EXIT
          END
        OTHERWISE                          /*  Give all other DB2 data*/
          DO                               /*  average performance   */
            SET &STORCLAS = 'STANDARD'
            EXIT
          END
```

Figure 106. Sample storage class ACS routine

13.4.5 Base configuration

The base SMS configuration defines the systems that will be managed as part of an SMSPLEX. Systems are defined either singly by name, or systems with a sysplex can be defined as a group using the name of the sysplex.

The base SMS configuration also specifies:

- A default management class applied to SMS data sets that are allocated without a management class.
- A default unit used for non-SMS allocations that do not specify a device type.
- A default device geometry specifying the number of bytes per track and tracks per cylinder. This is used for space-calculation for data sets that are allocated without a specific unit request.

13.4.6 Control data sets

DFSMSdfp uses the following control data sets to maintain storage management policy:

- **Source control data set (SCDS)** — This contains an SMS configuration, which defines a storage management policy. You can define any number of SMS configurations, each of which has its own SCDS. Then, you select one SMS configuration to be the installation storage management policy and make an active working copy of it in an ACDS.
- **Active control data set (ACDS)** — When you activate an SCDS, its contents are copied to an ACDS. The current ACDS contains a copy of the most recently activated configuration. All systems in an SMS complex use this configuration to manage storage. You can define any number of SCDSs, but only one can be put in the ACDS.
- **Communication data set (COMMDS)** — This serves as the primary means of SMS communication among systems in the SMS complex. The active systems in an SMS complex access the COMMDS for current SMS complex information.

DFSMShsm maintains an inventory of the contents of the volumes that it owns in a set of control data sets.

- **Migration control data set (MCDS)** — The MCDS contains information about the migrated user or application data sets.
- **Backup control data set (BCDS)** — The BCDS contains information about the backup copies of user or application data sets.

- **Offline control data set (OCDS)** — The OCDS contains information about the data sets on migration and backup tape volumes used by DFSMShsm.

- **Journal data set (JRNL)** — It is a sequential data set used by DFSMShsm to log all updates to the DFSMShsm CDSs. It facilitates recovery of a CDS in case of errors.

The first phase of automatic backup is the backing up of the DFSMShsm control data sets (CDSs). This process is called CDS version backup. It is controlled by HSM parameters.

CDS version backup backs up all the CDS data sets and the JRNL data set. The data sets can be backed up using either IDCAMS or concurrent copy function with DFSMSdss which significantly reduces the duration of the backup process. This reduction increases the availability of DFSMShsm functions which are not available while CDS backup is occurring.

Using the CDSVERSIONBACKUP parameter, you can specify how many versions of the CDS backups be kept active. Each time CDS version backup runs, DFSMShsm assigns a new version number to the newly-created backup data sets and uses that version number as the final qualifier of the CDS backup data set names. CDS backups can be directed to either tape or DASD. DFSMShsm always catalogs the backup copies that it creates.

13.4.7 Logs and journal

DFSMShsm maintains the following two kinds of logs for storage administrators to track its activities.

The **DFSMShsm log data sets** provide information about events on a particular processing unit and about commands entered with the LOG command. The DFSMShsm log records this information in chronological order.

The **activity logs** report on the backup, dump, migration, ABARS, and command processing of DFSMShsm in your system. DFSMShsm has five activity logs:

- The migration activity log provides information about space management activity, including MIGRATE commands for volumes and levels, interval migration, and automatic primary and automatic secondary space management.

- The backup activity log provides information about automatic backup, volume command backup, and volume command recovery activities.

- The dump activity log provides information about automatic dump, command volume dump, and command volume restore activities.

- The ABARS activity log provides information about aggregate backup and recovery activities.

- The command activity log provides information about TAPECOPY and TAPEREPL activity, and records error messages or informational messages that occur during low-level internal service processing.

13.5 User interface

DFSMS provides a common user interface for storage administrators to manage their policies.

13.5.1 Interactive Storage Management Facility

The Interactive Storage Management Facility(ISMF) is used as an interface for the administration of other DFSMS components. ISMF allows you to:

- Manage data sets
- Manage volumes, including DASD, tape and optical
- Manage SMS constructs
- Manage SMS ACS routines
- Manage SMS base configuration
- Manage tape and optical libraries
- Manage aggregate groups
- Submit IDCAMS DCOLLECT jobs
- Access DFSMSrmm ISPF dialog
- Access NaviQuest

ISMF presents data to users and storage administrators initially as lists, and provides functions that allow lists to be tailored by sorting or suppressing entries. When lists have been created, the results can be saved and input to ISMF again. Once objects have been identified from a list, ISMF allows functions to be performed against them.

The following screen shows the ISMF primary option menu.

```
Panel  Help
------------------------------------------------------------------------------
                    ISMF PRIMARY OPTION MENU - DFSMS/MVS 1.5
Enter Selection or Command ===>

Select one of the following options and press Enter:

0  ISMF Profile              - Specify ISMF User Profile
1  Data Set                  - Perform Functions Against Data Sets
2  Volume                    - Perform Functions Against Volumes
3  Management Class          - Specify Data Set Backup and Migration Criteria
4  Data Class                - Specify Data Set Allocation Parameters
5  Storage Class             - Specify Data Set Performance and Availability
6  Storage Group             - Specify Volume Names and Free Space Thresholds
7  Automatic Class Selection - Specify ACS Routines and Test Criteria
8  Control Data Set          - Specify System Names and Default Criteria
9  Aggregate Group           - Specify Data Set Recovery Parameters
10 Library Management        - Specify Library and Drive Configurations
11 Enhanced ACS Management   - Perform Enhanced Test/Configuration Management
C  Data Collection           - Process Data Collection Function
L  List                      - Perform Functions Against Saved ISMF Lists
R  Removable Media Manager   - Perform Functions Against Removable Media
 F1=Help    F2=Split   F3=End     F4=Return  F7=Up      F8=Down    F9=Swap
F10=Left    F11=Right  F12=Cursor
```

13.5.2 NaviQuest

DFSMS NaviQuest is an extension to ISMF that helps implement, test, and
verify the SMS environment. NaviQuest is installed under the ISMF Primary
Option Menu (Option 11). NaviQuest is an integrated component of
DFSMSdfp since DFSMS 1.4.

With NaviQuest you can:

- Create and execute DFSMS implementation test cases either on-line or in
  batch, and compare the results against a set of previous or expected
  results.

- Use NaviQuest in the initial design and testing phase of your first
  automatic class selection (ACS) routines and SMS configuration.

- Run selected DFSMS testing functions in batch.

- Define, alter, or display:

    - A data class
    - A storage class
    - An aggregate group
    - An SMS base configuration

- Display a management class

- Define or alter a VIO, Dummy, Object, or Object Backup storage group

- Run jobs in batch to alter or define pool or tape storage groups.

- Create reports interactively or in batch.

- Use NaviQuest reports during your DFSMS planning and design, for both DASD and for tape data migration.

- Perform ongoing storage administration activities.

- Use NaviQuest for production assurance prior to any changes to your DFSMS environment. The cross-reference facilities simplify the tasks of testing and verifying SMS configuration changes.

The following screen shows the NaviQuest primary option menu.

```
 Panel  Help
 ------------------------------------------------------------------------------
           ENHANCED ACS MANAGEMENT - NaviQuest PRIMARY OPTION MENU
 Enter Selection or Command ===>

 Select one of the following options and press Enter:

   1   Test Case Generation
   2   ACS Test Listings Comparison
   3   Enhanced ACS Test Listing
   4   Test Case Update with Test Results
   5   SMS Report Generation
   6   Model Commands Generation
   7   Batch Testing/Configuration Management
   X   Exit




 F1=Help     F2=Split   F3=End     F4=Return  F7=Up      F8=Down    F9=Swap
 F10=Left    F11=Right  F12=Cursor
```

13.6 Scheduling

DFSMShsm automatically creates backup versions or migrate data sets on specified days, beginning at a specified time of day.

The data sets must meet eligibility criteria defined in the storage management policy. They must be on SMS-managed or DFSMShsm-managed volumes that have been designated for automatic functions.

You can schedule the following DFSMShsm automatic functions:

- **Automatic Backup (AB)** — Automatic backup is performed in four phases:

  1. Backing up the control data sets at the beginning

  2. Moving data sets that have been backed up by command from ML1 volumes to backup volumes

  3. Backing up migrated data sets that are in need of backup

  4. Backing up all changed data sets except those in certain management classes

- **Automatic Dump (AD)** — Dump processing is performed in three phases. These phases are:

  1. Deleting expired dump copies

  2. Dumping all DFSMShsm-managed volumes and system-managed volumes for which automatic dump is requested

  3. Deleting and uncataloging excess dump VTOC copy data sets

- **Primary Space Management (PSM)** — PSM stops processing if it is able to reduce the volumes' occupancy below the required occupancy threshold, or when no further data sets are eligible for processing. PSM performs the following tasks in the specified order:

  1. Automatically deletes expired data sets.

  2. Removes unused space from data sets, by compressing PDS data sets and releasing over-allocated space.

  3. Moves data sets from DFSMShsm-managed volumes to DFSMShsm-owned volumes to provide a specified amount of free space on each primary volume.

- **Secondary Space Management (SSM)** — SSM is usually scheduled to occur before PSM. However, it performs administrative tasks rather than moving data from the primary volumes. SSM automatically cleans up ML1 volumes and the migration control data set (MCDS) by:

    - Scratching expired data sets.

    - Deleting records that are no longer required from the MCDS.

    - Automatically migrating data sets from migration level 1 volumes to migration level 2 volumes if they have reached their residence limit on ML1. Data sets on ML1 that require a backup are flagged as such and are not migrated until the backup occurs.

Interval Migration (IM) — IM is a subset of the tasks of primary space management, and is automatically scheduled on an hourly basis.

The following example shows how to schedule primary space management. For other automatic functions, see the *DFSMS/MVS V1R5 DFSMShsm Storage Administration Guide*, SH21-1076.

To start primary space management on Saturday and Wednesday, you can specify:

```
DEFINE PRIMARYSPMGMTCYCLE(YNNNYNN) -

CYCLESTARTDATE(99/08/07)
```

To start primary space management at 2 a.m. on the days defined in the above example, you can specify:

```
SETSYS PRIMARYSPMGMTSTART(0200 0400)
```

Primary space management would not start at 2 a.m. if, for example, at that time the system was not operating or automatic migration was held. The PRIMARYSPMGMTSTART parameter of the SETSYS command controls the processing window of automatic primary space management.

13.7 Removable Media Manager

DFSMSrmm provides policy-driven management for tape data sets and media. Storage administrators define policies that govern the retention and vaulting of volumes. DFSMSrmm not only manages the installation's tape data sets and volumes, but also controls the shelves where volumes reside, in all locations except for automated tape library dataservers.

13.7.1 Overview

DFSMSrmm is an online media management system designed to manage an installation's system-managed and non-system-managed tape media, including 3420 reels and 3480 and 3490 cartridge system tapes.

DFSMSrmm, IBM removable media management, is available as a DFSMS functional component and as an optional feature of OS/390.

DFSMSrmm manages your removable media resources, including cartridge tapes and reels, by providing these functions:

- **Storage management** — DFSMSrmm provides you with the opportunity to begin automating your installation-defined tape management policies, much as you do now with DFSMS management classes.

  DFSMSrmm provides utility programs designed to automate many tasks, such as initializing tape, extraction, expiration processing, and vault policy management processing.

  The DFSMSrmm ISPF dialog and TSO subcommands provide an easy interface to the functions available. Users require different levels of authorization for DFSMSrmm functions, and can be limited to the access they require using RACF FACILITY class profiles.

- **Library management** — You can create tape libraries, or collections of tape media associated with tape drives, to balance the work of your tape drives and operators.

  DFSMSrmm can manage these system-managed tape libraries:

  - The IBM 3494 Automated Tape Library Dataservers and IBM 3494 Automated Tape Library Dataservers, including contained Virtual Tape Servers

  - The IBM 3495 Automated Tape Library Dataserver

  - The Manual IBM 3495 Tape Library Dataserver Model M10

  The IBM 3494 Tape Library Dataserver supports 3490E and 3590 tape subsystems. The automated models of the IBM 3495 Tape Library Dataserver support 3490, 3490E, and 3590 tape subsystems. The manual models of the IBM 3495 Tape Library Dataserver support 3490 and 3490E tape subsystems.

  DFSMSrmm can also manage non-system-managed or traditional tape libraries, including automated libraries which are not system-managed, such as the IBM 3494 and IBM 3495 Tape Library Dataservers under Basic Tape Library Support (BTLS) control.

- **Shelf management** — DFSMSrmm groups information about removable media by shelves into a central online inventory, and keeps track of the volumes residing on those shelves.

- **Volume management** — DFSMSrmm helps to manage the movement and retention of tape volumes throughout their life cycle.

  DFSMSrmm helps you manage the movement of removable media between different locations.

- **Data set management** — DFSMSrmm records information about the data sets on the tape volumes to validate volume and data set information and to help maintain data integrity. It can also control the retention and movement of those data sets and volumes.

  DFSMSrmm provides full 44-character data set name validation for tape data sets.

13.7.2 Key features

DFSMSrmm provides following functions:

- Central control for the following resources related to the removable media in an enterprise:
    - Data sets
    - Owners of data sets, volumes, and policies
    - Retention and movement policies known as vital record specifications (VRS)
    - Software products
    - Shelf and storage locations where volumes reside
    - Volumes

- Expanded customer defined storage locations — Policies can be defined using almost any eight-character, or less, installation selected storage location name value.

- Retention in location until expiration — Retention and movement policy processing is extended by an additional policy that retains data in a location only until it is expired.

- Manual move control — Allows volumes to be removed from automated movement using the VRS, so that they can be held for special processing or can be manually directed.

- JOBNAME option for retention and movement policies — VRS can include a JOBNAME operand that can specify the names of jobs to which the policy applies.

- Retention date and VRS policy at the data set level to provide simple identification of how retention is determined.

- Ability to search on the LOAN location.

- Support for separate specification of default retention information for data sets left open or which were closed by abend processing.

- Utilities to convert to DFRMM are included with the product.

- An option for pool definitions that allows pools to be named, so you can use the name instead of the existing pool prefix on operator messages and displays.

- An option for defining the minimum retention period for data sets controlled through catalog retention, to allow time for the data set to be created and cataloged.

- Policy management of data sets and volumes, where you can:
  - Specify multiple retention and movement criteria in separate retention and movement policies
  - Override the volume expiration date by defining a policy
  - Use the DFSMSrmm vital record specification to override the expiration date so that the vital record specification controls all retention and movement a specified number of days beyond the retention date

- Provide an Application Programming Interface (API) to add, delete, change, list, and search information for users with appropriate authority

- Support requests for multi-volume and multi-data set chains as a single TSO subcommand or API request

- Support separate retention and vaulting policies making the conversion from other products easier

- Support environments with shared and non-shared catalogs removing the need to check the catalog status of all recorded data sets other than on an exception basis

- Enable special handling instructions through a special DD statement pointing to a data set containing the information needed by the tape operator

See the *DFSMS/MVS DFSMSrmm Guide and Reference*, SC26-4931 for more details.

13.8 Other functions

The following sections describe the services and facilities provided by DFSMS that are part of the storage management process.

13.8.1 Performance management

DFSMSdfp provides performance management by selectively managing the use of cache controller resources and selection devices that can meet performance requirements at allocation. Dynamic Cache Management Extended (DCME) dynamically determines the use of cache in a storage subsystem for each data set when it is accessed.

DFSMSdfp also stripes data sets to improve sequential access performance based on their storage class attributes. The latest release, DFSMS R10, allows you to stripe VSAM data sets in addition to traditional sequential data sets.

13.8.2 Tape Mount Management

Tape Mount Management (TMM) is a proven technique of SMS implementation that enables you to increase tape capacity usage and reduce tape mount activities at the same time. TMM is suitable for data sets that are relatively small compared to the capacity of tape media, and that are not required to be taken offsite.

TMM is implemented by using SMS ACS routines to redirect the allocation of selected tape data sets to DASD. TMM initially involves moving data in the reverse of the normal storage hierarchy. Eligible tape data sets are initially re-directed to DASD rather than being allowed to move to tape. This DASD is used as a buffer to collect a large number of these data sets. DFSMShsm is then used to sweep this buffer and move the relocated data sets to DFSMShsm-owned migration volumes.

This process both reduces the total number of tape mounts that are required, since DFSMShsm fills its tape volumes, and restricts the periods during the DFSMShsm sweep when the few remaining tape mounts occur. This allows other work to be scheduled around these periods.

You can use the volume mount analyzer (VMA) to study tape mount activity, monitor tape media use, and implement tape mount management at your installation. The volume mount analyzer produces reports that you can use to tailor data classes, management classes, and automatic class selection (ACS) routine filters, depending on your specific tape usage profile. Using tape mount management, you can maximize the use of your tape media and reduce your tape mounts.

See *DFSMS/MVS V1R5 Using the Volume Mount Analyzer*, SC26-4925 for more details.

13.8.3 DFSMS Optimizer

DFSMS Optimizer is a storage resource manager. Rather than manage storage directly, it reports on and automates other components of DFSMS.

The DFSMS Optimizer can:

- Analyze I/O subsystem performance at the subsystem, device, or data set level.
- Identify the proper level in the storage hierarchy for data.
- Determine the proper policies for data sets.
- Assess the cost and value of the current storage management policy applied to each data set.
- Analyze cost recovery of systems management.
- Automatically tune and monitor critical DFSMShsm system functions.
- Manage multiple DFSMShsm tasks running either within a sysplex or in multiple unrelated systems using the HSM Monitor.

DFSMS Optimizer has two principle components, the **SMS Optimizer**, which provides reporting for SMS components, and the **HSM Monitor/Tuner** (see Figure 107).
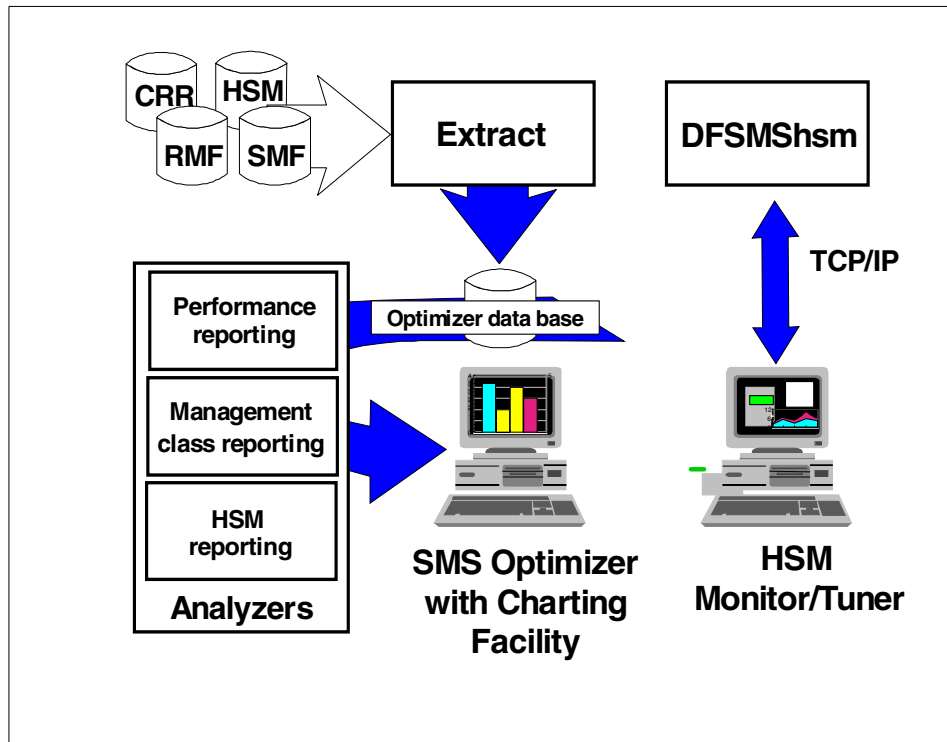
Figure 107. DFSMS Optimizer components

13.8.3.1 SMS Optimizer

The SMS Optimizer uses input data from several sources such as Cache RMF Reporter (CRR), Resource Measurement Facility (RMF), System Management Facility (SMF), and DFSMShsm CDS records. It processes them using an extract program that merges the data and builds the Optimizer database. The Optimizer data base consists of a number of sequential data sets.

By specifying different filters, you can produce reports that help you build a detailed storage management picture of your enterprise. The reports are run as batch jobs, and produce three separate types of output:

- Tabulated results
- A comma-separated variable (CSV) file that can be imported into a spreadsheet

- A chart file that can be downloaded to a workstation and the contents displayed graphically using the **DFSMS Optimizer Workstation Charting Facility**. This is a Java application that automatically makes color charts from the Optimizer report data through its own GUI.

The SMS Optimizer produces performance and management class reports. Performance management reports focus on:

- Current state analysis, including:
    - Reporting on average holding time and cache hit ratios
    - Reporting on NVS usage and subsystem lower interface utilization
    - Sorting top data sets, volumes and subsystems by I/O rate, MB transferred, I/O response time, and I/O intensity.
- Batch Window Analysis:
    - Identifies jobs and data sets that are candidates for striping or data in memory
    - Identifies which jobs are I/O bound
    - Identifies data sets with a high percentage of service time
    - Determines performance cost components
    - Calculates daily MB cost of cache and NVS storage
- What-if simulations, identifying what effects various changes such as the following might have:
    - Changing the size of cache storage
    - CPU hardware compression
    - Data set striping
    - Conversion to extended format data sets

The management class analyzer provides cost-benefit analysis and what-if simulations of management class policies for SMS and non-SMS data. With this analyzer, you use your storage costs and data access patterns as input when you run a report. You can also use the analyzer to predict the amount of level 0 (L0) and migration level 1 (ML1) storage required if you use a new set of migration attributes.

Additionally, you can use the Management Class Analyzer to determine the most cost-effective management class assignments for data sets you want to convert to SMS, or to tune your current management class settings to reduce possible DFSMShsm or DFHSM thrashing.

The DFSMS Optimizer Version 1, Release 2 enables you to monitor your data management activities in a Java-based environment under OS/2, Windows 95, and Windows NT.

13.8.3.2 HSM Monitor/Tuner

You can use the HSM Monitor/Tuner to keep track of multiple DFSMShsm systems through an on-line TCP/IP connection. You can also play back a log file previously created by a host DFSMS Optimizer HSM Monitor/Tuner session. The HSM Monitor/Tuner provides real-time monitoring of HSM automatic migration, back-up, and dump functions; recall and processing statistics; and management work elements. You can also issue HSM commands, cancel HSM tasks, and modify the trigger thresholds that control HSM activity right from your workstation.

In addition, the HSM Monitor/Tuner can automatically monitor and tune DFHSM and DFSMShsm through the use of more than 150 trigger events that invoke REXX automation routines. You can customize these REXX execs to issue commands directly to the host that triggered the event to issue NetView alerts or other automated responses as corrective actions.

See 14.5, "StorWatch DFSMShsm Monitor" on page 292 for additional information.

13.8.4 Tape library management

Tape libraries that are to be SMS managed need to be defined in the SMS configuration using ISMF option 10. When you define tape libraries, you need to specify the following:

- **LIBRARY ID** specifies the hardware ID associated with the tape library being defined. This is the value returned in response to a Read Device Characteristics command; it is defined by the customer engineer at the time of the library installation.

- **CONSOLE NAME** specifies the name of the MVS console associated with the tape library being defined.

- **ENTRY DEFAULT DATA CLASS** specifies the data class to be used to determine the interchange attributes for private volumes entered into the tape library. No other attributes of the data class are used. You can use the cartridge entry installation exit (CBRUXENT) to assign appropriate interchange values that override the default data class specifications.

- **ENTRY DEFAULT USE ATTRIBUTE** specifies the default volume use attribute for the cartridges entered into the library being defined. You must choose between PRIVATE or SCRATCH.

- **EJECT DEFAULT** specifies the default action for the TCDB volume record when a tape cartridge is ejected from the library being defined. You must choose between PURGE, which will cause the volume record to be deleted from the TCDB on ejection of the tape, or KEEP, which will retain the record in the TCDB.

- **SCRATCH THRESHOLD** specifies the minimum acceptable number of scratch volumes for each media type in the library being defined.

- **INITIAL ONLINE STATUS** specifies how the library being defined will be associated to each system or system group in the SMS complex each time this SCDS is activated. Valid values include on-line, off-line, or not connected.

13.9 Specifications

In this section, we present the specification for installing DFSMS.

13.9.1 Prerequisites

DFSMS R10 is delivered with OS/390 Version 2 Release 10 or later. Customers with OS/390 Version 2 Release 5 or 6 are not able to order DFSMS R10 separately, since this release is now tightly packaged into the OS/390 operating system.

The following programs or equivalents are prerequisites:

- High Level Assembler (HLA) for MVS, VM, and VSE (5696-234) with current maintenance

- MVS/ESA SP (TM) Version 5 Release 2.0, JES2 (5655-068) or JES3 (5655-069) and higher, or OS/390 version 2 (5647-A01) release 5 or 6 with their supported versions of JES2 and JES3

- System Modification Program Extended (SMP/E) Version 1 Release 8 Modification level 1 (5668-949) for installation and maintenance

13.9.2 Priced features

Priced features available to DFSMS are as follows:

- DFSMS Optimizer feature — DFSMS Optimizer Release 2 is available as either a feature of DFSMS/MVS V1R2 (or higher), or as a standalone product, the DFSMS Optimizer for MVS/ESA and OS/390 (program number 5655-OPT).

> **Note**
>
> OS/390 Network File System (OS/390 NFS), formerly delivered as DFSMS/MVS Network File System, has previously been available as part of DFSMS/MVS and was also available as a nonexclusive element of the OS/390 product. Beginning with the OS/390 Version 2 Release 6, NFS is only available on OS/390 Version 2.

Chapter 14. StorWatch

This chapter covers the StorWatch family of products that are available to help you configure, monitor, and manage your storage environment.

14.1 Overview

StorWatch — IBM's Enterprise Storage Resource Management (ESRM) solution — is a growing software family whose goal is to enable storage administrators to efficiently manage storage resources from any location within an enterprise. Widely dispersed, disparate storage resources will ultimately be viewed and managed through a single, cohesive control point.

14.1.1 Key management areas of enterprise storage resources

IBM has identified eight key management areas that need to be addressed by any comprehensive enterprise storage resource management solution. Although existing software solutions might individually address many of these areas, relatively few solutions support multiple areas. If functions to address these areas are combined into a single solution, storage administrators will be able to manage distributed storage resources much more effectively. As a result, organizations should receive a better return on their investment in storage equipment through better utilization, availability, and performance — all of which helps reduce the overall cost of storage ownership.

The StorWatch family of products addresses the following storage management areas:

- **Asset management** — Identifying storage resources by parameters that include type, model, serial number, features, location, acquisition date, price, and maintenance information.

- **Capacity management** — Viewing common capacity metrics (total capacity, utilization percentage, utilization trends, fragmentation, and efficiency) across a variety of storage subsystems.

- **Configuration management** — Viewing and modifying hardware configurations for host adapters, caches, and storage devices at the operating system, software application, and licensed internal code levels.

- **Data, device, and media migration** — Simplifying the movement of data to new device technologies. Performing outboard data movement and replication within a storage subsystem.

- **Event and alert management** — Monitoring events and receiving notification of exception conditions such as threshold, initiating exception handlers, and logging information.

- **Performance management** — Monitoring current and historical performance information such as path, device, and cache utilization and I/O operations per second. Simulating "what if" situations to help predict performance impact.

- **Policy management** — Ensuring that enterprise-wide storage management policies are consistently followed. Propagating policies to all platforms and locations from a central repository.

- **Removable media management** — Tracking tape and optical media volumes, and sharing tape and optical devices among applications.

14.1.2 Product family

The StorWatch product family is designed with a consistent look-and-feel for each of the products which have browser-based user interfaces. This consistency of the user interface can help storage administrators decrease the required learning time for products which perform the many aspects of enterprise storage resource management. The StorWatch family is growing supporting new storage solutions as they become available.

Today, the StorWatch family is composed of the following components:

- StorWatch Reporter
- StorWatch Enterprise Storage Server Expert
- StorWatch Serial Storage Expert
- StorWatch Enterprise Storage Server Specialist
- StorWatch SAN Data Gateway Specialist
- StorWatch SAN Data Gateway S20 Specialist
- StorWatch Fibre Channel Switch Specialist
- StorWatch Fibre Channel RAID Specialist
- StorWatch DFSMShsm Monitor

14.2 StorWatch Expert

The StorWatch Expert is an optional product. Its purpose is asset management, capacity management, and performance management for the ESS and VTS.

This section explains the role of these products:

- Asset management

- Capacity management
- Performance management

Asset management

The asset management functions of the ESS Expert provide you with an overview, by server, of serial numbers, vital product data, and other asset relevant information.

The ESS Expert also provides charts with trends and projections of ESS storage server growth.

Capacity management

If you want to know how much storage each server has allocated on ESSs, the ESS Expert can tell you. The ESS Expert has a summary of hosts and their storage on ESS systems, or you can request a summary for each ESS.

Performance management

The ESS Expert also can tell you how the ESS performs. While S/390 customers have many tools available to monitor device performance, these kind of tools are rare in the open systems world. Therefore, this kind of information might be of special interest for open systems users.

Particularly, the ESS Expert provides the following information:

- Number of I/O requests
- Number of bytes transferred
- Read and write response time
- Cache use statistics

14.3 StorWatch Specialists

This section details the current — and always growing — set of StorWatch Specialists.

14.3.1 StorWatch Enterprise Storage Server Specialist

The StorWatch ESS Specialist comes with the ESS product. The ESS Specialist must be used to configure an ESS. Apart from the configuration function for ESS, the ESS Specialist can be used to administer Copy Services functions to set up FlashCopy T0 copies or Peer-to-Peer Remote Copy. It allows storage administrators to remotely monitor and manage ESS. Using a Java 1.1-compliant internet browser — such as Netscape Communicator or Microsoft Internet Explorer — storage administrators can access StorWatch Enterprise Storage Specialist from work, home, or on the

road through secure network connections, to gain additional control and flexibility in managing ESS.

See 7.1, "The IBM Enterprise Storage Server (ESS)" on page 121. Also, see *IBM Enterprise Storage Server,* SG24-5465 for additional information on ESS and ESS Specialist.

14.3.2 StorWatch SAN Data Gateway Specialist

The StorWatch SAN Data Gateway Specialist provides an easy to use graphical interface to configure, manage, and service the Gateway. The tools enable status monitoring, and dynamic changes to the Gateway setup, providing the added convenience of centralized management of multiple Gateways. See 9.3, "SAN Data Gateway for UltraSCSI disk and tape storage" on page 189 for detail.

14.3.3 StorWatch SAN Data Gateway S20 Specialist

The StorWatch Specialist has two functions — the SAN Data Gateway S20 Manager, and the Service Utility. The Manager provides the tools for configuring and managing multiple Gateways across the enterprise. The capability exists to concatenate up to 16 physical disks into a single composite drive with the StorWatch Specialist.

Establishing sets of mirrored RAID 1 data (up to three copies), management of the mirrors, and re-synchronizing as necessary are all handled through the StorWatch SAN Data Gateway S20 Specialist (splitting off a copy of a mirrored group is currently limited to Sun, HP, Compaq — AlphaServers — and non-clustered NT Servers). The service functions enable status monitoring, error reporting and recovery, and dynamic changes to the Gateway setup. See 9.1, "SAN Data Gateway for serial disk" on page 185 for detail.

14.3.4 StorWatch Fibre Channel Switch Specialist

The IBM StorWatch SAN Fibre Channel Switch Specialist provides a comprehensive set of management tools that support a Web browser interface for flexible, easy-to-use integration into existing enterprise storage management structures. The Specialist provides security and data integrity by limiting (zoning) host system attachment to specific storage systems and devices. See Chapter 3, "Storage Area Network Initiative" on page 19 for detail.

14.4 StorWatch Serial Storage Expert

IBM StorWatch Serial Storage Expert is a software tool to help increase the productivity of storage administrators who manage the configuration and status of SSA disk networks.

StorWatch consists of two parts — a storage network Planner and a storage network Live Viewer.

- The Planner can be used to plan new SSA storage networks from scratch or to plan modifications to storage networks that have already been installed. The Planner has a palette of storage network components from which you can choose disk drives, enclosures, adapters and connections. Planning a storage network is as simple as selecting a component, dropping it on the canvas and connecting the components together. To work with existing networks, load a Management Set created by the Live Viewer into the Planner.

- The Live Viewer discovers and monitors SSA storage networks that have already been installed through TCP/IP connections to the SSA host computers. Simply enter the names of the host computers of your SSA networks to discover all of the SSA devices attached to the hosts and draw a physical representation of the storage network. IBM's StorWatch Serial Storage Expert will monitor the SSA devices for faults and for topology changes and notify you of any changes.

IBM StorWatch Serial Storage Expert simplifies SSA network topologies with support for SSA enclosures (7133, 7131, and 3527). Users can select and place enclosure icons on the IBM StorWatch Serial Storage Expert canvas and install SSA disk drives into the enclosure. IBM StorWatch Serial Storage Expert remembers disk placement across Storage Expert sessions by saving location information in the Management Set. Enclosure icons have the same features and properties as real SSA enclosures. Connections between drives are automatically resolved into internal enclosure or physical cables between enclosures or enclosures and adapters.

You can further simplify your canvas by using the Storage Network Navigator. The navigator is a separate window that contains an icon for each storage network in the open Management Set. From the navigator window, you can choose to hide or show the network on the canvas by first selecting the corresponding network icon and then clicking on the appropriate button. The Storage Network Navigator displays the state of the devices in the network by changing the look of the network icon, even if the network is hidden. Administrators can choose to monitor the Storage Network Navigator instead

of the canvas to monitor network status and show only those networks that are experiencing problems.

IBM StorWatch Serial Storage Expert improves the time required to discover SSA storage networks and events by up to two times as compared to StorX. This increase in speed makes it even more useful as a day-to-day storage network monitor. Storage Expert supports SSA adapters and disk drives attached to Intel platform servers running Microsoft (TM) Windows NT 4.0. Install the Windows NT agent on any server you wish to discover and add the server IP address to the Management Set. You can have both AIX and Windows NT storage networks in the same Management Set at the same time.

A report function provides a method of extracting detailed information about the SSA devices within the Management Set. You can export most of the collected attributes into a flat file format suitable for importing into popular spreadsheet or database programs, allowing you to create your own custom reports with your favorite tools.

Events are saved with the Management Set, providing a history of the devices in the Management Set. Storage Expert will automatically navigate to a resource identified by an event when the user clicks on the resource name in the event text. Users can copy selected event text to the clipboard and export the entire log to a file name specified by the user via a file dialog.

See *Monitoring and Managing IBM SSA Disk Subsystems*, SG24-5251 for more details.

14.5 StorWatch DFSMShsm Monitor

The StorWatch DFSMShsm Monitor provides the complete HSM monitoring and tuning function of the HSM Monitor/Tuner component of the DFSMS Optimizer. It incorporates all of the functional enhancements that were made to the HSM Monitor/Tuner in Version 1 Release 2 of the Optimizer, including the following:

- Improved stability and reliability
- Better event recording and monitoring
- Expanded user interface
- Enhanced triggering and automation
- Command response at the workstation
- Selected HSM task termination

Also, the monitor workstation components can now run on Windows 95 and Windows NT, as well as OS/2.

The monitor has been separated into two components to improve both reliability and stability. The monitoring function has been put in its own MVS address space, while the workstation communications take place independently from monitoring activity. The monitoring of HSM is no longer driven by requests for information from the workstation. The monitoring occurs through timed gathering of information and through support added to HSM. This enhances the information obtained by the monitor.

By keeping the monitoring function separate, workstations can be allocated to display HSM activity as needed, and a workstation session is no longer required to keep the monitor active. Even if no workstations are active, the monitor continues to:

- Gather information
- Record the information in the monitor log file
- Invoke the monitor trigger automation routines

The monitor log file can be accessed for playback by the workstation to review prior HSM activity. The monitor logs no longer need to be on the same volume and can be SMS managed.

New information is gathered by the monitor for the DFSMShsm automatic functions of primary space management, secondary space management, interval migrations, automatic backup and automatic dump.

New panels have been added to the workstation user interface to allow the viewing of the automatic function information, and the volumes and data sets the automatic functions are processing.

A new feature allows the termination of the current volume or data set being processed by a DFSMShsm automatic function. Following the termination, the automatic function will continue processing with the next volume or data set in the queue.

The triggering events have been greatly expanded to coincide with the gathering of the new information. Triggers are called when certain events occur in the course of automatic function processing. Furthermore, the REXX automation function that is invoked when a trigger occurs has been moved to the host monitor component to improve availability and reliability. Automation occurs even if no workstation is active.

The workstation code for the monitor has been rewritten in Java and can be run using Windows 95, Windows NT, or OS/2. Two or more workstations can be run concurrently in different operating system environments.

The monitor has the capability to issue DFSMShsm commands from the workstation and receive the responses to these commands in a scrollable window at the workstation.

See 13.8.3, "DFSMS Optimizer" on page 280. Also, see *DFSMS Optimizer: The New HSM Monitor/Tuner,* SG24-5248 for more information.

Chapter 15. Tivoli Storage Area Network Solutions

In this chapter we describe various storage management products other than TSM available Tivoli Systems. Those products include:

- Tivoli SANergy
- Tivoli Decision Support
- Tivoli Storage Network Manager
- Tivoli Storage Resource Manager

Tivoli Storage Area Network (SAN) Solutions provide a comprehensive Open Systems SAN solution built upon a scalable architecture to allow customers to cost-effectively manage new and emerging SAN technologies. Tivoli's SAN Solutions enable customers to fully exploit even the most complex SAN configurations, by offering the SAN management and exploitation capabilities to deploy, allocate, maintain, protect, and share SAN resources (see Figure 108).
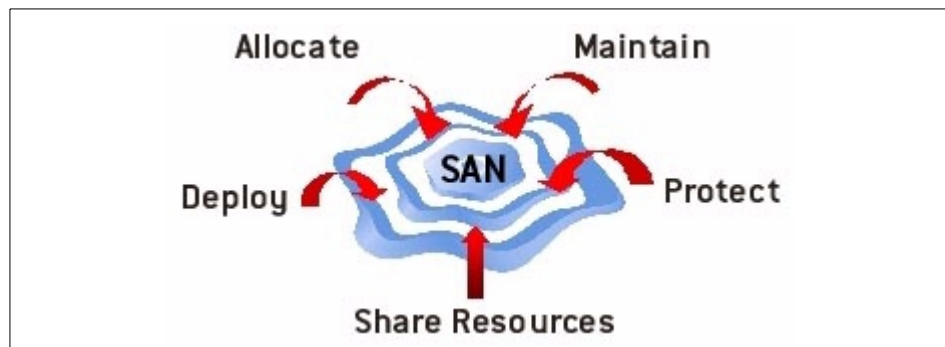
Figure 108. Tivoli Storage Area Network (SAN) Solutions

15.1 Tivoli Storage Network Manager

Tivoli Storage Network Manager is designed to facilitate the management of a SAN, from top to bottom, irrespective of vendor software and hardware platforms. In order to accomplish this, we first need to examine some of the new emerging technologies from Tivoli, such as Tivoli Kernel Services and Tivoli Presentation Services, before we can fully expose Tivoli Storage Network Manager. Figure 109 presents a holistic view of Tivoli Storage Network Manager and a SAN:
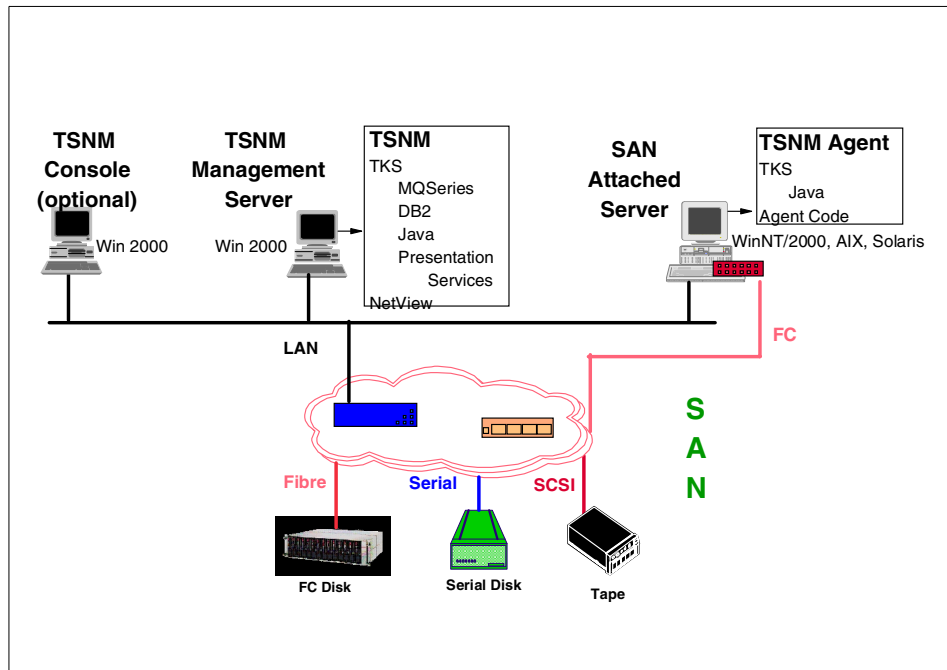
Figure 109. TSNM — The Big Picture

15.1.1 Overview

Tivoli Storage Network Manager provides the interface for managing your Storage Area Network infrastructure. Tivoli Storage Network Manager is a scalable solution utilizing open SAN standards. This comprehensive solution discovers, allocates, monitors, automates, and manages SAN fabric components and attached disk storage resources.

Benefits:

- Minimize costly inefficiencies
- Minimize downtime
- Sustain dependable information availability
- Reduce administrative workloads

15.1.2 Key features

In this section we present key features of Tivoli SAN Solutions.

15.1.2.1 Network management

This section covers the network management features.

- SAN topology discovery

  This feature performs SAN topology and disk resource discovery, which provides a visual confirmation of initial SAN architecture and element connectivity. Supports standards-compliant vendors' FC components.

- Continuous monitoring

  This feature provides continuous montitoring, problem identification and resolution techniques to assist in the maintainability of the SAN infrastructure, so that applications can continue to have access to data. Administrators can launch, from within Tivoli Storage Network Manager, the specific SAN component management software to assist in closure of problems, and to perform device configuration and management functions.

- Single interface

  When used with Tivoli NetView, you can monitor and control your SAN infrastructure and devices from the same console that you use to manage your LAN and WAN.

15.1.2.2 Disk management
This feature easily and securely allocates storage resources to the appropriate host systems for application processing (at the LUN level).

15.1.2.3 File system policy management
Policy driven threshold management automatically allocates and extends additional disk resources and file systems to maintain continuous application processing, reducing administrative workload.

15.1.2.4 Event reporting
SAN-related events are forwarded to SNMP (Simple Network Management Protocol) and/or Tivoli Event Console (TEC) to report on SAN activities to the designated administrator or management console.

15.1.3 Specifications
The Tivoli Storage Network Manager Server is supported on Windows 2000. The managed host platforms are supported on Windows NTR, Windows 2000, IBM AIXR, and Sun Solaris. Visit Tivoli Storage Network Manager Requirements for specific platform version or editions, and the most complete and up-to-date requirements and support information.

Summarized in Table 9 are the general requirements in order to run Tivoli Storage Network Manager. The console and manager can be run on the same machine or on separate machines.

Table 9. Tivoli Storage Manager: MVS and Windows NT server examples

Environment to manage	Tivoli Storage Manager features
MVS server with a 2-rack IBM 3494 library	Tivoli Storage Manager for MVS Base Server NETWORK SPACEMGMT ADVANCED DEVICES
Windows NT server with an IBM 3570 library	Tivoli Storage Manager for Windows NT Base Server NETWORK
18 Windows NT clients	18 CLIENTS
10 Netware clients	10 CLIENTS
6 AIX clients with HSM managed by the MVS server	6 CLIENTS
2 Sun Solaris clients	2 CLIENTS
3 HP-UX clients	3 CLIENTS
5 Lotus Notes on Windows NT	5 Tivoli Data Protection for Lotus Notes on Windows NT
3 Oracle clients on AIX	3 Tivoli Data Protection for Oracle on AIX
3 AIX SAP/R3 databases with Oracle	3 CLIENTS 3 BACKINT for Oracle

Console Manager Managed Host:

The following minimum requirements must be met:

- Machine Minimum Pentium III 400 Minimum Pentium III 600
- Minimum Disk Space Required to Run see the product README file see the product README file see the product README file
- Operating System Windows 2000 Professional, Server, or Advanced Server Editions Windows 2000 Advanced Server Edition Windows NT 4.0

Server, Windows 2000 Professional, Server, or Advanced Editions, AIXR 4.3.3, and Solaris? 2.7

- Memory Minimum 256MB Minimum 1 GB Minimum 100MB

Devices Supported:

Tivoli Storage Network Manager supports standards-compliant vendor FC components.

15.2 Tivoli SANergy

In this section we cover Tivoli SANergy.

15.2.1 Overview

Tivoli SANergy is software that runs on the computers connected to a SAN:

- It greatly simplifies SAN-storage administration through the power of sharing.
- It extends industry-standard networking to utilize the Gigabit bandwidth of the SAN.
- It broadens increased-data-availability protection to include any application.
- It can reduce the total amount of storage required in a SAN.

Tivoli SANergy eliminates the one-to-one relationship between the number of SAN-connected computers and the number of disk volumes needed by those computers. Tivoli SANergy transparently enables multiple computers to share single disk volumes on the SAN-storage. In fact, it allows many combinations of computers running Windows NT, Windows 2000, MacOS, Irix, Solaris, AIX, Tru64, Red Hat Linux, and DG/UX to all share the same disk volumes at the same time across platforms. And if the applications running on those computers are capable, Tivoli SANergy even enables the transparent sharing of the same files at the same time across platforms

15.2.2 Key features

Tivoli SANergy enables single disk volumes to be shared between multiple computers:

- It reduces the total number of disk volumes that need to be built and maintained in a SAN.

- It reduces the time between volume rebuilds (reformats) and reassignments because of the reduction in the number of volumes and the increase in the size of those volumes.

- It allows disk space unused by one computer to be used by any other computer sharing the same disk volume.

- It eliminates wasted disk storage space.

- It enables single file systems to be shared between computers using different operating systems.

- It enables single files to be shared between multiple computers

- It enables single files to be shared between computers using different operating systems

15.2.3 What's new in 2.2

Note: The following paragraphs were taken from the Tivoli Web site.

New features in Tivoli SANergy 2.2:

- Performance enhancements

  These are enhancements to software that significantly improve the shared file performance for such applications as multi-server Web hosting and other environments whose shared I/O demands are focused on read-only file access.

- Published APIs for partners

  Two Tivoli SANergy APIs are now available to partners, enabling them to integrate with and leverage Tivoli SANergy by creating new MDC platforms and supported file systems. One API allows File System vendors to enable a SANergy MDC to share access to their volumes, thereby extending the standard SANergy support for built-in file systems (NTFS on NT/W2k, UFS on Solaris, and EXT2 on Linux). The other API enables companies such as NAS server vendors to add MDC support for SANergy clients to their products, affording the opportunity for NAS servers to provide faster and more efficient file sharing to their clients. These two APIs will be made available via the Tivoli Ready? program, and are in addition to the existing SNMP interfaces to SANergy for administration and configuration.

- Linux support

  Offers a completely new and powerful way for Linux environments to leverage SANS. Full support is offered for computers running Red Hat Linux, as both a Tivoli SANergy Meta Data Controller (MDC) and as a Tivoli SANergy client. In addition, Tivoli SANergy also adds support for systems running DG/UX.

- Expanded client support

  SANergy MDCs on UNIX platforms now fully support Windows NT, Windows 2000 and Macintosh clients, as has the Windows NT/2000-based MDCs in prior releases.

- Enhanced Macintosh support

  Offers improved functionality, performance and device support for MacOS 8 and 9. As noted above, Macintosh clients are now also supported by all Tivoli SANergy MDCs.

Part 4. Implementing enterprise storage administration

Chapter 16. Enterprise storage administration

Enterprises are being forced to optimize their use of resources to enable them to compete in today's business environment. Storage management is not an exception to this trend.

Before you can begin to optimize the use of your storage resources, you need to understand the storage environment that exists within your enterprise today.

In particular, you need to be able to answer the following questions:

- Do you fully understand how your enterprise is managing data today?

  This must include data management across all the enterprise, not just the data currently on the mainframe. If the current storage-management processes are not understood, then the enterprise is at risk of wasting or losing resources.

- What are your current data management costs?

  The key to making cost-effective business decisions is understanding how data is currently managed, and the costs associated with this.

- Can you predict the impact of change to the storage management environment, and in particular changes in storage management policies?

These questions can be answered in a more reliable way, by implementing an enterprise storage administration.

16.1 Understanding enterprise storage administration

There are many potential benefits for enterprise storage administration at an enterprise level.

Although the demand for storage is growing, often companies do not use the full capacity of existing devices. Or, they are forced to acquire additional storage devices in differing formats, when they do in fact have the capacity they require for growth, but in an incompatible format.

Today's customers main concerns with data management are:

- **Multivendor environments** — Most enterprises contain many different systems, supplied by different vendors, running incompatible software.

- **Distributed storage environments** — These environments grow rapidly, and their tools must evolve also. In a truly distributed environment, which

could be Web-enabled, it is also difficult to create time windows to process traditional data management housekeeping tasks.

- **Manual inventory** — Keeping an inventory of backups today is a manual paper-based process for many systems.
- **Labor intensive tasks** — Manually taking backups and storing tapes in a safe place are usually labor-intensive tasks, and are subject to daily operation pressures, leading to errors. Recovery of backups performed by users may also be more labor intensive and time consuming than they prefer. Business decisions still need to be made, even during the recovery process. This demands balancing the speed and completeness of the full recovery against the time taken to perform a recovery to the point of failure.
- **Inconsistent or inadequate data management** — Data distributed across a number of servers is more difficult to back up and to restore to a consistent point in time, in an acceptable deadline.
- **Management driven by exceptions** — Management by exception becomes aware of issues once a problem has occurred, so it is usually too late to make better operation decisions to prevent the problem.

16.1.1 Overview

The enterprise storage administration is a set of requirements that defines storage performance, security, and availability. The administration uses storage management policies to automatically manage the enterprise storage devices (direct access, tape, and optical), and to automate the enterprise's data backup.

A dedicated group of persons, called the *storage administration group*, manages the storage resources of your installation. They are a part of the data processing center, and provide superior data processing services to all levels of users, while maintaining cost efficiency.

Inside the storage administration group, the storage administrator is initially responsible for the planning and implementation of the environment, and later, responsible for storage planning, storage availability, and technical tasks.

In today's environment, storage administrators have developed automated storage management processes, so that the labor-intensive nature of storage management is reduced. Although the daily, tedious storage management tasks are reduced, complex planning and problem-resolution skills are still required.

Storage administrators base their plans and tasks on service-level agreements (SLAs). These SLAs document expectations between the storage administration group and other groups, such as the user groups.

To help users understand storage management processes and specify their requirements, a user's guide is needed; such a guide must document all policies and procedures that apply to users. The guide should be oriented to the users' requirements in terms of ease of use, language, and updates, so they can easily apply it to their needs.

16.1.2 Benefits of enterprise storage administration

In a world where the individual purchases may be small, the solution to storage management has often been to spend through the problems.

Many reasons explain that situation:

- The cost of a new hard disk for a workstation is relatively low, and the cost of new disks for all employees of a company may never be clearly seen.
- The time taken to reinstall all the software and data to the disk is usually not taken into account.
- The fact that the original disk was littered with files that were not essential, and were referenced only infrequently, is completely overlooked.

Analysis shows that the actual cost of obtaining and physically maintaining storage media (hardware, software and repairs) is less than one third of the total cost spent on storage management.

Statistical surveys have shown that managing data on a storage system costs, each year, seven times the device purchase price. Therefore, automating the management of these data by an enterprise storage administration enables a gain of at least 30%.

If storage could be managed more effectively, significant savings could be made. Most of these benefits can be reached by assigning storage management tasks and responsibilities to a specialized storage administration group. In that way, you alleviate the redundancy of effort that occurs when users are required to manage their own storage. This specialized group can perform these functions more effectively and remove the burden of storage management from the users.

Figure 110 shows that before centralizing data management, the different platforms are managed differently, with different standards and reliability, and replicated storage. After centralizing data management, all the data are

managed with the same standard reliability by a storage management group, and the local disk expansion is controlled.
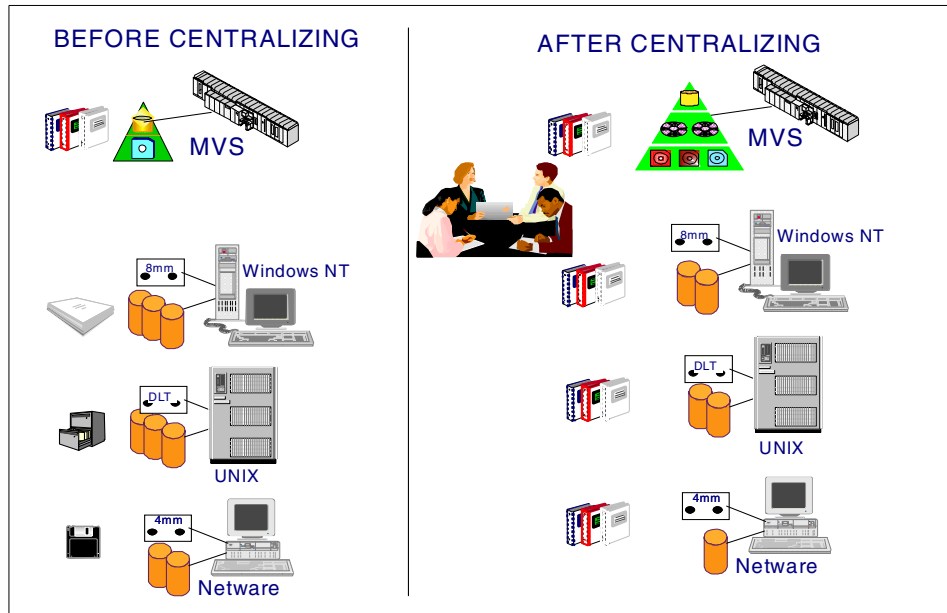
Figure 110. Before-and-after centralizing storage administration

A centralized storage management with enterprise-wide scope introduces many new enhancements:

- **Improved user services** — Improved user services result from service-level agreements, which allow both service providers and users to understand the minimum levels of services that have been mutually agreed upon.

  The storage administration group:

  - Negotiates and maintains service-level agreements with user groups, to define and document realistic requirements for space, data retention, and management of individual data sets and files

  - Meets user requirements for performance, space, availability, and improved response times

  - Provides a focal point within the enterprise for solutions to problems and improved procedures concerning storage management

  - Provides a standard interface to users for all storage functions

- Eliminates the disruption to your users when installing future storage devices

- Allows users to focus on their primary responsibilities

• **Coordination of resources** — The storage administration group is better able to view the data processing system from an enterprise-wide perspective. System integrity can be more efficiently maintained with centralized monitoring. This is particularly important where resources are shared across multiple heterogeneous platforms. In these cases, there is an absolute need for a single point of management control, to:

- Standardize enterprise-wide procedures for performance, backup and recovery, and security

- Use the full capacity of your devices

- Match the logical needs of your user's data to the physical characteristics of storage devices

• **Improved data processes** — The enterprise storage administration improves the data processes. Many verifications can be transferred to the centralized storage administration:

- Ensuring that data is backed up according to application requirements

- Ensuring that space and resources are available when needed

- Automating more storage management tasks

- Establishing centralized control of storage resources

- Tracking the data process while accessing to the central storage devices

- Automating many of the current storage administration procedures

• **Resource authorization policies** — The need for controlling access to data and system resources is increasing in most installations. Providing effective access authorization involves:

- Centralizing the management of the authorization and security in your installation

- Tailoring authorization and security procedures to your organization's needs

- Monitoring and controlling access to data

- Automating authorization and monitoring procedures

• **Reduced costs** — With technologies where multiple platforms can share the same devices, centralizing administration can significantly reduce the costs, by:

- Increasing productivity by separating users from the details of performance, availability, and space management

- Avoiding the duplication of efforts that occurs when users, applications developers, and system administrators all manage their own storage

- Controlling new hardware acquisitions by using existing resources more efficiently

- Controlling additional staffing, which is otherwise associated with manual storage management

- **Ongoing education programs** — A comprehensive education program is an important factor in ensuring a successful data processing facility. The storage administration group with responsibility for enterprise-wide implementation of new methods can coordinate education efforts among the various groups: users, operators, application developers, and management.

16.2 Identifying storage administration tasks

Three major task areas are key to the centralized approach to enterprise storage management:

- **Managing data and data availability** — Your storage administration group provides user services throughout the life cycle of a file or data set. Storage management software provides a way to manage data at each stage in the life cycle, by using a product such as Tivoli Storage Manager and/or DFSMS/MVS, to assign management classes. However, in a nonautomated environment, storage administrators still need to perform certain basic functions manually.

- **Managing storage pools** — You can match the logical needs of data to the physical characteristics of storage devices. Users will no longer need to know the details of their hardware configuration, such as device characteristics, volume serial numbers, or space limitations.

  Storage pools make users more productive while increasing the effectiveness of space, of availability, and of performance management. Another benefit of storage groups is the reduction of the disruptive effects of device installation.

- **Configuring storage subsystems** — Your selection of storage devices, and your configuration of DASD strings and optical devices, directly affects your storage administration group's ability to meet space, performance, and availability requirements.

Consider the performance and availability needs of the data in each of your storage groups, as you determine your hardware configuration. For example, storage classes that specify high performance require high-performance hardware. Plan to cooperate closely with other technical support areas to balance your resources and configure the storage subsystem.

Managing data sets and objects, managing storage pools, and configuring the storage subsystem are interrelated. Together, they help you satisfy your installation's needs in terms of performance, availability, reliability, security, and use of space.

These three task areas, as shown in Figure 111, can be broken into the following tasks:

- 16.2.1, "Service-level management and control" on page 312
- 16.2.2, "Technology evaluation" on page 314
- 16.2.3, "Capacity planning and monitoring" on page 315
- 16.2.4, "I/O performance management" on page 316
- 16.2.5, "Establishing standards" on page 316
- 16.2.6, "Technical support" on page 317
- 16.2.7, "Disaster recovery planning" on page 318
- 16.2.8, "Device installation and migration" on page 320
- 16.2.9, "Storage administration tools development" on page 320
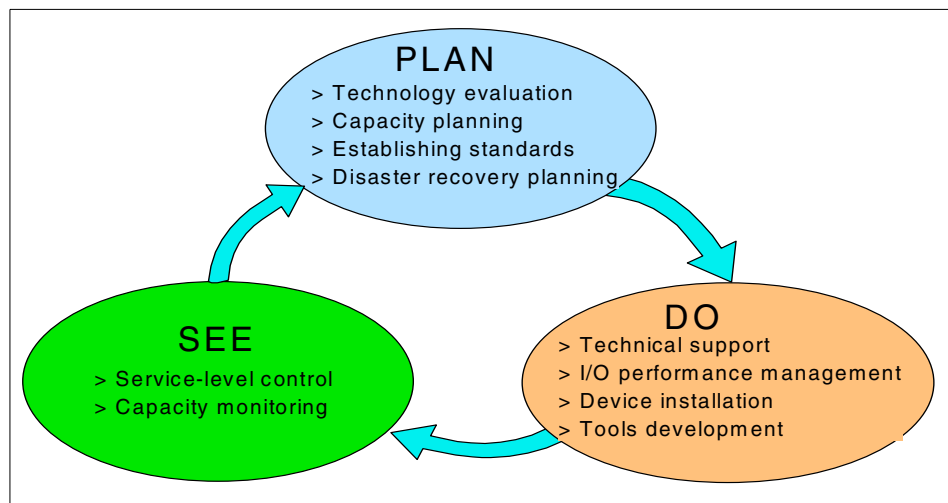
Figure 111. Storage administration tasks

It is possible that today in your enterprise, these tasks are not performed by storage administrators. However, if these tasks are being performed with no input from storage administrators, then results may not be predictable.

16.2.1 Service-level management and control

The storage administration group and the user groups negotiate the service-level agreements until the contract is mutually acceptable. It should have clearly measurable terms, so that everyone knows when the agreements are met. Finally, the management of each group should participate in the approval of the agreements.

16.2.1.1 Service-level agreement with user groups

The centralization of storage management responsibilities can change the role of the users in your installation. When you define and maintain your storage environment, the users need to be able to describe their data in terms of required service levels; they will need to accurately specify logical requirements for space, performance, and availability.

With service-level agreements, user groups are committed to follow the policies and conventions that the storage administration group has established to provide those services.

Remember that your users may also be service providers to others, and that their requirements are driven by the needs of their own users. However, ultimately, the level of service that you can provide will be determined by the value that the enterprise places on the data in question.

16.2.1.2 Service-level agreement with operation groups

Some of the agreements you reach with the user community depend on the ability of other groups in the data processing center, to provide reliable service. For example, your ability to meet performance criteria can depend on the network planning group.

Therefore, negotiate a separate service-level agreement with the operations groups. This agreement should clearly document their responsibilities in meeting service levels, and helps ensure successful completion of storage management jobs, and full backup and recovery services.

16.2.1.3 Contents of service-level agreement

A service-level agreement should contain:

- **Procedures and responsibilities for**:
    - Providing service requirements

- Reviewing and updating service-level agreements
- Measuring, reporting, and reviewing service levels
- How improvements in service levels are to be accomplished
- How service levels and delivery are to be charged back to the users
- What user actions can nullify or make void a service level — for example, failure to comply with installation standards
- What on-line operator's guide documents the maintenance and recovery procedures

- **Objectives; what improvements are planned for**:
  - Response time
  - Availability, capacity, and reliability
  - Turn-around time for problem resolution
  - Continuity of service during changes

- **Scope of the agreement**:
  - What devices and types of storage media are to be covered by the storage group (for example, disk, optical, tape, and control units)
  - Which systems, applications, and services the service-level agreement covers
  - Exceptions. For example, some data require alternative backup mechanisms or special placement, or may be specifically excluded from the agreement (such as temporary files, or data that has already been archived)
  - What plan period and duration the service-level agreement covers; for example, a 1-year period from July 1, 1999 through June 30, 2000, with service levels measured and reported monthly, and with the agreements reviewed and updated at every 6 months

16.2.1.4 Reporting on service-level measurements

Establish an effective means of measuring and reporting the levels and quality of your services. Incorporate your measurements into the storage administration group's regular responsibilities, and include the following:

- Measuring the free space available in storage groups at the start of the different processing time periods during the day
- Counting the number of failures due to insufficient space or unavailable resource

- Counting the time waiting for resources (storage place, media, communication)

- Tracking the number and types of storage device failures

- Measuring the amount of space reclaimed from periodic deletion of expired data and the migration of inactive data

- Charting the actual space used, versus planned space, by user account or by application

16.2.2 Technology evaluation

As vendors develop and market new technology, you must determine which of the competing technologies, if any, will best meet the needs of your enterprise. When you are managing storage on an enterprise basis, technology evaluation becomes even more complex, and you have also to decide which is the most appropriate platform for data placement.

As with service-level agreements, technology evaluation cannot take place in isolation. Storage administrators need to confer with other groups before making fixed plans.

Technology evaluation should consider, not just the function and performance of new devices, but also such factors as:

- Costs for purchasing and maintaining, as well as the costs associated with having a device unavailable due to servicing

- The physical resources that a device may require for operation

- Software technologies, determining which are the products that best meet your needs for managing the storage environment

- Synergies that may exist between hardware and software

- The probable life expectancy of the technology

If, as the end result of a technology review, a decision is made in favor of a new purchase, a formal request for purchase will need to be made to your finance department, for more hardware or higher levels of software.

A business case provides the detailed requirements that support the request for purchase, such as hard disk capacity requirements and software requirements. It can include a comparison of profitability and effectiveness of a proposed system and the current one. The request for purchase and the business case are usually combined into a single document.

16.2.3 Capacity planning and monitoring

Capacity planning is perhaps the most important of the storage management disciplines. If you cannot provide the storage capacity required for the business, then business essential processes will not occur, and the enterprise will lose money. On the other hand, acquiring excess capacity means that the enterprise has spent money on an unused resource.

Capacity planning cannot be carried out by any one group in isolation. An effective capacity plan will be produced only after many diverse groups have input their requirements.

In an enterprise where storage may not have traditionally been centrally managed, user groups may have been used to planning for and purchasing on their own behalf, and of seeing storage devices as their personal property, rather than as an enterprise resource. In an enterprise-managed environment, this loss of ownership may become a difficult issue to deal with, during capacity planning discussions, particularly during the transition period.

Application developers and business analysts provide the raw input for capacity planning, determining the number of files, estimating the amount of data that will be processed, as well as the rate that transactions will be arriving.

However, storage administrators need to add to this process. They are best positioned to be able to assess things like:

- The amount of additional space required for managing files

- The effective capacity of devices; some pools may be able to tolerate higher levels of occupancy than others

- How space is to be provided: by reclaiming storage volumes and devices, by purchasing new storage devices, or by a combination of both

- What the impact would be of additional equipment on the installation's environmental requirements, such as floor space, power, and air conditioning

- The impact of time- and resource-consuming requests

- The performance implications

Because storage administrators are monitoring the performance of systems continuously, they are also best placed to identify problems such as slow creep in capacity required, or the growth that you have when no growth should have occurred.

16.2.4 I/O performance management

I/O performance management ensures that access to data is sustained at acceptable levels. If performance does not meet user requirements, business is impacted either by staff being less productive while waiting for system response, or worse by customers being unable to obtain service.

I/O performance management can be linked to frequency of access. Data that is accessed infrequently usually does not need such rapid response as data that is used very regularly. These decisions govern the type of media that the data should be stored on, and to some extent, its location. High-frequency, low-response-time data should be stored on fast disk devices; lower frequency, higher response time data may be stored on optical disks. Information accessed only rarely can be stored on tape.

Performance management should also include calculating the actual costs of the storage management functions that are performed. Costing should include elements such as the cost of CPU processing, media costs per unit, the costs of any manual intervention that is required, and the costs for media mounts and movement. These costs should be used when considering changes to storage management policies.

In addition, many other factors may influence performance, such as network, but the storage administrator cannot take engagements in that area. He or she can just make service-level agreements with other groups.

16.2.5 Establishing standards

Determining standards for the storage environment is one of the most critical tasks for a storage administration group. However, standards cannot be determined in isolation. A standard that does not complement the current business practices of your organization may be more costly than having no standard at all.

To justify its existence, a standard must make data processing or storage administration more cost effective for the enterprise. For example, a standard for data set names may be justified by allowing easy identification of the data owned by one application. It will potentially allow filtering on data set names to be used, which could have the following impacts:

- Identify this data as part of a particular application, and hence, also identify the owner of this data — for example, by identifying an archive package.

- Simplify security administration — for example, by the creation of RACF generic profiles for these data sets.

- Make it easier to identify data that needs to be backed up together with Tivoli Storage Manager, or DFSMSdss, or DFSMShsm ABARS filters.
- Allow grouping of application data sets in unique user catalogs.
- Ease the coding of SMS ACS routines that determine the location and management of this data.

Having determined standards that should be obeyed, you must ensure that these standards are actually being implemented, including:

- Restricting use of data set or file names not conforming to installation standards.
- Restricting use of unsupported access methods.
- Ensuring that data allocation standards are being met.

Standards also need to be flexible enough to allow for change to be incorporated without invalidating the current standards. Changing standards is a costly process, not only because of the amount of rework that is required to bring existing structures into line with the new process, but also because errors are likely to be introduced during the transition period, when more than one standard exists.

16.2.6 Technical support

Technical support involves maintaining the storage environment in a state where all agreed user requirements are being met. Technical support personnel can also develop good working relationships with other groups, providing them problem resolution, operation advice and education.

16.2.6.1 Second-level support

Technical support provides second-level support to a Help Desk and develops the skills of operators and users.

Operators need to understand the storage environments and how to use commands correctly during problem diagnosis and resolution. This is especially true in an environment with multiple heterogeneous systems. Operations need to be aware of the differences between systems, as well as the linkages between systems. Include operators in your education program, both to ease their entry to the new environments and to ensure their compliance with future changes.

The second-level support includes:

- Providing second-level coverage for storage problems reported to the Help Desk. As with the first level, the Help Desk answers the calls from users.
- Providing advice on application backup and recovery procedures.
- Providing samples and tools to allow standardized calls to storage services from applications.
- Providing storage management education for operations as needed.

16.2.6.2 User assistance
Although some users may already be experienced in storage management, others may need an introduction to the storage management policies and practices in your installation.

Consider spending time with user group representatives to look at their groups' data and applications, to help them understand their responsibilities and the implications of various service options (for example, backup frequency), and to develop education programs suited to their needs.

Users also need to understand and adhere to allocation standards and naming conventions. Teach users to start thinking of their storage needs in terms of data, number and format of records, rather than space or volumes. They are likely to cooperate more fully with these policies if you provide education programs that explain the reasons behind the standards.

The major challenge is to simplify the storage management environment as much as possible for the users. The fewer decisions that they are required to make, the fewer are the mistakes that are possible.

It is therefore important to develop a user's guide that describes in detail the facilities available to the users, including the operator commands that control processing and system procedures. Your user's guide should be a continually evolving document that reflects changes in the storage environment, services offered, and responsibilities. Ideally, the guide should be on-line for all users, eliminating producing, maintaining, and distributing more than one copy.

16.2.7 Disaster recovery planning
The storage administrator is responsible for the integrity and availability of the computer data. The enterprise must be able to survive a catastrophic failure, such as fire or sabotage, affecting its data processing equipment. The storage administration group must be involved in the design, planning, or implementation of a disaster recovery solution for your enterprise.

You need to be able to prepare your data processing environment so that it can be recovered in the event of a disaster. To do this, you will need to perform the following:

1. Determine the requirements of your business processes, by performing a risk analysis and a business impact analysis.

2. Determine the data processing requirements. These include for each business process:

    • Required recovery time

    • Maximum allowable data loss

    • Computing power required to run and disk storage required

    • Dependence on other applications or data

3. Design the recovery and backup solution. (This should be a high-level design.)

4. Evaluate and purchase the resources required to implement the desired solution.

5. Implement the solution.

6. Maintain and test your recovery solution.

Accomplishing these tasks will involve a complete cycle of analysis, design, and implementation of disaster recovery. Some organizations may choose not to perform all the work related to planning for and implementing disaster recovery themselves, preferring to use specialized experts for this. Whatever your installation decides, the storage administrators are deeply involved in this process.

Many enterprises also invest in disaster avoidance, as well as planning measures of data redundancy to take in case a real disaster occurs.

The minimum preventive measures usually consist of good on-site recovery procedures, to avoid having routine problems escalate into a disaster. This will require, for instance:

  • Spare hardware
  • Regular backups
  • Skilled operations staff
  • Building fortification
  • Improved fire protection
  • Rigorous access control.

The time and cost involved in implementing and maintaining such measures can be justified through comparison with the time and cost involved in dealing with the situation at the time of a disaster.

If, in spite of the preventive measures, a potentially disastrous incident occurs, you will have to determine if it requires onsite recovery or disaster recovery procedures. The situation is not necessarily obvious, so the decision criteria have to be carefully prepared, as part of planning and implementing a disaster recovery solution.

Disaster recovery is a complex topic, and there are a number of redbooks devoted entirely to it. For example, *DFSMShsm ABARS and Mainstar Solution*, SG24-5089

16.2.8 Device installation and migration

Changes in data types and business procedures can eventually change some of your requirements for specific storage groups. You might fit the storage environment:

- Work with your hardware installation group and systems programming group to ensure nondisruptive device installation.
- Advise on hardware configuration definition for new storage devices.
- Define and maintain pool definitions as required.
- Add or delete volumes from storage pools as required.
- Initialize devices if required.
- Populate new devices without causing application disruption.

After the device installation, the storage administration group is responsible for allocating storage resources. The group must be seen as an honest broker in allocating these resources, making decisions that benefit the business. If a user has the financial justification for a resource, the storage administrator needs also to be able to supply these resources.

16.2.9 Storage administration tools development

You will need to put in place a number of tools and processes, to enable you to:

- Maintain and report on the storage management environment.
- Define interfaces for the users.
- Define interfaces for the storage administration operators.

As the storage environment changes, so may these tools and processes.

Even if you are implementing tools supplied by a vendor, you will need to tailor them to meet your installations standards. This may involve putting in place batch processes and procedures to monitor and report on the functioning of your storage environment. It may also involve installing real-time monitors that enable administrators or operators to have a current view of the performance of the storage environment.

The output from the storage management tools that you install and develop will be input to all other storage processes. It is important that these tools provide data that is both reliable and relevant.

16.2.9.1 Tools in the MVS environment

In the MVS environment, tools that can be used to generate raw data for further processing include:

- **DCOLLECT** — The access method services (IDCAMS) DCOLLECT command provides the ability to collect information related to active data sets, migrated data sets, backup data sets, volumes, DASD capacity planning, and capacity planning for DFSMShsm-owned tapes. The information can be collected from DASD or DFSMShsm-owned tape.

- **System Management Facilities (SMF)** — SMF provides information at both the data set level (for example I/O statistics) and at the device level (RMF records including information about response time and service times). Most storage products will log data to SMF to document their performance, such as DFSMShsm statistics records and DFSMSrmm security/audit records.

- **DFSMSrmm** — You can use the DFSMSrmm extract file to produce reports that document the current status of tape volumes and data sets and actions, that may be outstanding against them and the locations where they are stored.

- **Environmental Record Editing and Printing (EREP)** — EREP generates reports that help you and IBM personnel diagnose hardware and software problems.

16.2.9.2 Tools in the Open environment

In the Open environment, several tools can be used, depending on the storage administration platform, such as scheduler, log surveys, batches.

However, Tivoli Storage Manager is a cross-platform storage solution, and the following specific Tivoli Storage Manager features help the administrator:

- **Accounting** — Records the amount of data and the processing times for each session,

- **Centralizing the events** — The events from the clients and from the server can be centralized and filtered on one focus point,

- **Tivoli Plus for Tivoli Storage Manager**— This tool integrates the Tivoli Storage Manager survey and automates the management tasks from the Tivoli desktop interface.

16.3 Converting to enterprise storage administration

Converting to enterprise storage administration involves the users, the storage administration group, and other groups providing services to users.

First, it is important to analyze the implications of the storage administration in your enterprise, in terms of costs, benefits, and user involvement.

16.3.1 Costs of setting up a storage administration group

The costs of setting up a storage administration group include initial setup and training, and education for group members. Generally, when establishing any new group or department, most costs are realized up front, during the planning phases.

Emphasize to your upper management that set-up costs are one-time costs, and that current personnel might already possess transferable skills. Include all your resource estimates for staffing and equipment.

Your organization may need fewer people to manage storage, but those people need a greater level of skill and will bear more responsibility. Remember that if your enterprise has not centralized storage administration today, many tasks are performed by others who are not storage professionals. During the implementation of enterprise storage management, the number of professionals in the storage group may certainly increase; however, you will more than compensate for this by freeing other people from tasks that they were originally burdened with.

You will need to sell the benefits to user group management and to departments that may never have seen fixed expenses charged for storage services and face the need to budget for these charges. This will be difficult. Complicating the situation even further is the problem that the current owners of equipment may not be aware of the cost of managing their equipment, as they see no direct charges for this. The costs are potentially all being absorbed as overhead within their departments.

You also need to be able to come up with a solution that:

- Allows for the transfer of assets and handling their depreciation
- Does not cripple the budget for your new department
- Does not leave the original purchaser paying for equipment, he/she no longer owns and may no longer be the sole user of

16.3.2 Explaining benefits of centrally managed storage for users

An important step in establishing successful enterprise storage management is to transfer control of storage volumes from users to the storage administration group. This is the key to using storage space more efficiently, providing vastly improved service to users.

Users who understand the reasons behind the change and perceive its rewards, are more likely to view the transfer of storage management responsibilities as a relief. However, those who have not been prepared for the change and do not see its benefits, may not welcome the transfer.

You can produce a capability statement that documents the types of tasks that the group can perform for users. In some cases, it may also be possible to include small job biographies of some of your staff, so that user groups can see that the people they will be talking to understand the users environments.

To convey your message effectively, establish a positive rapport with users and devise an education program to show how the transfer of storage control enables users to do their jobs better, by decreasing the effort they must apply toward managing storage, while simultaneously increasing the performance of the storage subsystems they use.

16.3.3 Phases of enterprise storage administration

The process will be long and complex, and the task is too great to see in one view, so it needs to be broken into phases.

We have chosen the following:

1. Business justification

   You need to find a leverage point to justify this piece of work, for instance, standardizing backup processing by installing Tivoli Storage Manager.

   You need to identify the resources that will be impacted by this change, and any new resources that may be required. These resources include hardware, software and people.

2. Obtaining business approval

   Upper management agrees with your assessments of the benefits that the business will receive. At this point, you also need to obtain a sponsor from upper management for the project. This sponsor will help you obtain the project resources you require.

3. Developing standards and methodology

   Define the standards that you intend to implement, and the way that you intend to implement them.

4. Choosing a pilot environment

   If possible choose an environment that has:
   - Implemented storage management to some extent
   - Dedicated storage management staff
   - Processes and procedures that do not depend on other systems
   - A typical environment.

5. Skills upgrade

   To provide outstanding service, the storage administration group must have considerable expertise in storage management, specialized skills that most users and applications developers do not have. Data processing growth trends show that performance, availability, and space management problems often result when unskilled personnel perform storage management. In an enterprise that may contain many types of systems, it is essential to have members in the group with both cross platform skills and deep in-depth knowledge of the platforms to be supported.

   Thus it is important to ensure that your group has the skills required for enterprise storage management at this point, or has a project in place to acquire them, either by absorbing skills from other sections of the enterprise or by training or purchasing expertise from outside the enterprise.

6. Transition

   During the transition phase, more of your enterprises data will become centrally managed. This is perhaps the most complicated part of the project, as this is the time of most rapid change, and where problems arising from confusion caused by change are most likely.

   The transition phase requires additional staffing and resources because of potential duplication of effort — for example, producing listings in two different formats (the new and the old) for comparison — and unfamiliarity with systems.

It may be helpful to have a dedicated transition team who investigate, evaluate, and document new environments. If you have the skills available, this team could also begin to implement your new standards on these systems.

7. Consolidation and standardization

   After a system has completed transition, the next step involves fully integrating this system into the processes and procedures of the group.

Chapter 17. Implementing the storage administration group

In this chapter, we discuss the practical details of how to implement enterprise storage management.

The first step in establishing a complete enterprise-managed storage environment is to justify establishing a storage administration group to upper management. The following sections will help you gather information to build your business case. With upper management's backing, you can develop the implementation plan and begin moving toward enterprise management.

17.1 Developing core plans

This section helps you identify the major plans and tasks for implementing enterprise storage management.

The important tasks for the storage administration group are to develop the mission statement, the storage management plans, and the service-level agreements. After these key plans have been defined, the storage administration group updates your installation's existing plans to incorporate system-managed storage.

If you already have a storage management process in place, you can compare it to the descriptions in this chapter and consider adjusting your plans.

The following sections describe the types of plans that you will need to prepare or have modified to include your group.

17.1.1 Mission statement

You should clearly define the scope and content of your group's mission. This mission statement will provide a basis for planning, establishing, and maintaining centralized storage management.

If you are building a new storage administration group, you will want to demonstrate your value to the business. Focus on a few immediate objectives to show quick, highly visible results. For example, to offer tangible results, define storage groups or standard data set allocation.

Define immediate goals as they relate to long-term objectives. Consider short-term objectives as milestones on the road to realizing your ultimate goals. Other factors that can influence a departmental mission statement include:

- Your department's role in the data processing organization
- Your installation's experience with centralized storage management
- The particular space problems in your installation

Your mission statement should detail storage management activities that contribute value to your business. Some of the following activities are appropriate:

- Planning for the conversion to a centrally-managed environment
- Planning for storage subsystem and capacity requirements

17.1.1.1 Overall responsibilities

Before implementing the enterprise storage administration group, the other groups need to understand the responsibilities of the storage group to its users and the group's requirements from its users.

The storage administration group is responsible for:

- Providing storage management services that increase user productivity

- Providing education and guidance to users on storage-related issues

- Making efficient use of current and future storage resources

- Pursuing techniques that provide required levels of service in the most cost-effective manner

- Implementing plans that are compatible with the overall direction of the installation

- Managing projects that relate to the storage subsystem

- Measuring and reporting device service problems and their resolution,

- Performing administration tasks, including security tasks and the checking of logs

- Performing technical services, such as acceptance testing and identification of technical requirements for improving storage

- Tuning storage system performance, including load balancing and subsystem configuration

- Performing service management and planning, including negotiating and creating service-level agreements with users

- Developing and managing service-level agreements

- Monitoring and analyzing usage and performance

- Developing and executing data recovery procedures

- Providing input to and negotiating the content of the plans

- Reviewing and maintaining storage management process documentation

User group representatives are responsible for:

- Communicating their needs and providing feedback regarding the storage management process
- Providing their requirements on a regular basis
- Attending regular service reviews
- Communicating status of problems between the storage administration group and the users
- Reviewing and approving change control plans
- Reporting on storage group project status
- Providing business justification of storage requirements
- Communicating with the storage administration group on behalf of the user community

17.1.1.2 Immediate goals

To implement the installation's storage management policies, the storage administration group pursues the following objectives:

- Obtain ownership of all storage resources.
- Implement user independence from physical storage.
- Implement automatic storage management techniques.
- Standardize data management policies.
- Using DFSMS, or ABARS, or Tivoli Storage Manager DRM functions for disaster recovery can:
    - Promote user acceptance and understanding of policies
    - Provide user education about system-managed storage
    - Promote storage management as a business process
    - Establish a secure environment for data assets
    - Provide service measurement and reporting

17.1.1.3 Long-range goals

The storage administration group supports the following long-range goals:

- Add users and new storage devices nondisruptively
- Expand the scope of tasks and responsibilities to include new technologies as they are implemented

- Implement storage management as a business process
- Improve the quality of services

17.1.2 Storage management strategic plans

Your enterprise's strategic plans are the enterprise's overall plan for marketing and business, and will have input from all areas of the business. It includes the strategic technical plan, which is a two- to five-year outlook for the enterprise's computing resources. From a storage viewpoint, the plans should be updated to include the following items:

- Processes and additional hardware and software
- New applications
- Changes in the way existing applications are developed and written

17.1.2.1 Requirements plan

This is a list of needs for storage that you gather from your user groups. It should also cover:

- Software and hardware upgrades required to meet a business need
- Enhancements to service levels, such as increased availability of data
- Requirements for all types of data, including data sets and objects
- A procedure for collecting subsequent requirements

17.1.2.2 Capacity planning

Ongoing storage subsystem capacity planning helps you determine future device needs, so you can avoid last-minute crisis acquisitions.

Trend analysis is an important part of capacity planning. Monitor the recall and mount rates to check performance trends and make the appropriate storage subsystem capacity plans.

When updating the capacity plan to incorporate enterprise storage management, you need to identify:

- Procedures:
  - What people, departments, and business areas are to be involved with capacity planning, and what their roles and responsibilities are.
  - Sources and methods for gathering capacity planning data. For example what tools to use, and which members of the installation's departments are to provide their capacity requirements.
  - How space availability and usage charges are to be charged back to the users.

- That last point can be a contentious issue, especially if space management software is being used, There are many different approaches to charging for migrated data. The two basic approaches are to either charge migrated data at the same rate as on-line data or to charge some discount for migrated data. Charging for both migrated and backed up data may also present additional difficulties during the transition phase, when media previously purchased by individual groups is still being used for their data. A possible solution for this would be to either reduce or waive charges for amounts of data below a threshold based on the capacity transferred to the storage group.

- Projections must be made based on the following:
    - How much space is currently allocated, used, and not allocated?
    - What space is required for the next plan period?
    - How are space requirements to be fulfilled? For example, what space can be reclaimed from existing storage group volumes, and what space will have to be provided by buying additional storage devices?

- Scope covered by the capacity planning:
    - Devices and types of storage media
    - Systems and applications
    - Plan period and duration

Notice that the capacity plan's scope depends closely upon the scope of the service-level agreement.

Regular planning discussions with user groups are essential to effective capacity analysis. Their input is needed to determine not only current space and performance requirements, but also any future plans that affect your predictions for storage hardware needs. Storage capacity planning is closely tied to capacity planning for the entire installation. It involves close work with other data processing center support groups, such as system developers, networks, plans, and controls.

17.1.2.3 Commitment
The commitment plan is raised as a response to the capacity plan. The commitment plan is a financially based plan and is based on capital and expense limits, that have been decided by the company's financial arm.

This plan describes the amounts of money available for groups to spend to run their business during the next planning period. Purchases outside of these limits would need to be reviewed with higher management. From the

commitment plan, the storage group can then begin allocating the resources that will be available for the next planning period.

17.1.2.4 Recovery

The disaster recovery plan outlines requirements to recover data, equipment, and business processes in the event of a physical disaster. In that perspective, a copy of the data has to be stored offsite. Include storage management considerations in your installation's disaster recovery plan.

You also need to consider the requirements of data recoverability, where the cause of the loss is not a disaster. Most data losses are directly due to user errors, rather than to any more spectacular cause. Data recoverability at this level is often an application concern, but the storage management team may be required to provide suitable tools for backups to allow this to occur.

17.1.3 Storage management tactical plans

Tactical plans are the short-term response to the interaction between strategic plans and the day-to--day operation of the business.

17.1.3.1 Procurement

Hardware and software plans describe ordering, purchasing, and installation plans for storage devices, and the software to support them. Storage hardware planning needs to be coordinated with the data processing center's hardware planning specialists. Coordinate software installation with the system programming specialists.

17.1.3.2 Implementation and integration

An integration plan defines the tasks associated with the newly installed hardware and software, and their integration into existing storage and operating subsystems.

Your integration plan should be defined or updated to include tasks associated with the integration and installation of hardware and software, on existing storage and operating subsystems. The responsibility for developing and maintaining the integration plan could be shared by the storage operators and one of the storage administrators, who helps with reporting information.

17.1.4 Service level agreements

The service plan consists of one or more service-level agreements.

You might establish a well-publicized Help Desk, a phone line, or an electronic message facility, whatever ensures the most timely, courteous

response, so users can make requests, register problems, and obtain information or interface with the processes that exist currently within your enterprise. In larger installations, the group should use enhanced electronic communication features, like *Lotus Notes*, to communicate and manage informations.

Monitor the effectiveness of communication and the workload it adds by tracking problems and requests, and quantify your response time. Recurring problems can indicate a need for either additional user education, or a change in storage management policy or procedure.

In an SMS-managed environment, the service levels of your users are based on the SMS environment that you established. The service-level agreement defines service levels associated with storage groups, and states the SMS classes that fulfill service requirements for user data. The same functions are provided by the Tivoli Storage Manager policy set definitions.

17.2 Developing group goals

Defining a number of distinct goals allows the members of the storage group to direct their activities in the most appropriate directions. Goals also allow members to measure their achievements against a set of defined targets. When defining specific group and personal goals, input from various group members may be useful.

17.2.1 Identify tasks that the group will perform

The storage administration group works with user group representatives to define storage administration goals, plans, and processes, such as change control and problem control.

The main storage management tasks are:

- **Availability management** — Availability management is ensuring that data is backed up at appropriate intervals, depending upon the business, in a manner compatible with a timely and usable restore.

  Thus, tasks required to ensure availability include:

  - Defining and maintaining management classes or other documentation used for data backups and archives

  - Defining and maintaining tasks and processes used for disaster backup and recovery of critical application data, including offsite retention

  - Verifying that data required to be taken offsite is being moved appropriately

- Managing incremental backup of data
- Managing volume and data recovery as required
- Testing recovery procedures regularly

- **Space management** — Space management is optimizing the usage of the storage resources available. This means both optimizing the use of storage media and ensuring that data is accessible and located appropriately in the storage hierarchy.

  For Tivoli Storage Manager, define management definitions, so that the bigger and older files are migrated. To improve the recalling of these files, ensure that they are migrated to the fastest storage device.

  Space management tasks include:

  - Monitoring space usage and ensuring that this is within agreed service levels (availability and performance)
  - Defining thresholds, depending on the increase rate of the data
  - Releasing allocated but unused space
  - Migrating inactive data
  - Deleting unneeded data

- **Configuration management**

  Managing the physical and logical configuration of storage media and of storage management software includes:

  - Measuring and reporting space usage for all storage groups
  - Ensuring that sufficient free space exists across storage groups to satisfy allocation requests
  - Adding or deleting volumes in storage groups as required
  - Analyzing tape usage
  - Identifying tape volumes for system management and for automation

- **Housekeeping** — Housekeeping tasks are usually done daily. Tools may be developed to automate these tasks.

  Housekeeping tasks include:

  - Daily monitoring of the correct execution of storage management subsystems and processes
  - Monitoring the availability for data on storage hardware resources, and performance, and report the accounting
  - Monitoring system catalogs

- Restricting use of absolute track and cylinder allocation
- Ensuring optimum use of all storage space in the installed base
- Highlighting potential problems and alert appropriate interfaces
- Managing problems and change control situations, and alert appropriate interface groups

- **Performance support** — Tasks associated with performance support include:
  - Managing data placement within the storage hierarchy
  - Monitoring service times (media and communication wait)
  - Analyzing performance data. This may include the data for subsystems, cache, nonvolatile storage, back paths, volumes and data sets
  - Isolating causes of performance problems such as device skew, I/O bottlenecks or periods of intense I/O for specific data
  - Having isolated the cause of the problem, developing a plan to correct or minimize it. This could involve moving data to resolve conflicts, reallocating data on a more appropriate device, or changing or improving the way that data is accessed (for instance, by adding disks controllers)
  - Verifying that you are not recalling data the day that it is migrated, the number of bytes and data sets migrated is less than the number recalled
  - Providing performance input to other groups, for example to users for requirements plan, to network and system groups for performance plans
  - Maintaining performance levels of storage devices and volumes
  - Communicating roles to other storage administration groups

- **Technical support** — Work with the technical and performance departments to specify requirements for system support for technical enhancements and performance requirements:
  - Developing and maintaining operations procedures for daily routine storage management processes. These procedures should have meaningful exit codes to enhance or tailor functions supplied by products.
  - Maintaining data classes, management classes, and storage classes

- Assessing and support the evaluation of new products, facilities, and operating system releases from technical support

- Providing Help Desk service

- Maintaining skill levels and product awareness of storage administration group members, and provide user group education. Inform the groups of strategic planning changes or updates.

- Developing an on-line user's guide and help facilities, accessible through a HTTP browser in the distributed environments

- Holding regular user forums for education and communication to discuss problems and concerns

- Developing and provide installation storage management standards

- Reviewing standards and guidelines regularly

- **Storage capacity planning** — Planning affects most of the tasks. However, it is important to dissociate the planning tasks from the daily housekeeping tasks:

  - Reviewing the service-level agreements periodically with users and review plans and requirements

  - Providing input on disk, and optical, and tape requirements for the installation capacity plan

  - Working with the network group to plan the data growth

  - Reviewing the disaster recovery plan

17.2.2 Identify main interactions with other groups

A primary responsibility of the storage administration group is to provide improved user services. To provide excellent service to your users, you must understand their requirements, have the resources to meet them, and maintain an effective level of communication between groups.

The following tasks help you identify user requirements and create a cooperative work environment:

- Identifying user group representatives for the major application areas
- Listening to user requirements
- Negotiating service-level agreements with user groups
- Establishing channels of communication for user service requests
- Providing an education program
- Providing a user's guide that documents all policies and procedures

Users will support the transition to centralized storage management if you take the time to understand who they are, how they use computer resources, and how they view storage management. When it comes to the satisfaction of user needs, there is no substitute for frequent interaction and open dialogue between your group and representatives of the user community.

17.2.3 Establish reporting processes and procedures

The storage administration group works with user group representatives and the data processing department, to agree on measurement and reporting methods. This monitoring tracks service delivery from the user's perspective, and the success of the storage administration group's internal procedures.

For example, your group decides which tools to use, how often to measure service levels, and what kinds of measurements to take, to increase internal efficiency and to provide increased storage performance for users.

To do this, it will help if you:

* Provide a central contact for the daily running management of the overall storage management subsystem, and process, and reporting.
* Ensure that an effective escalation process is in place and documentation is provided to reflect relevant communication interfaces.
* Establish regular, standardized measurement and reporting procedures.
* Establish a method for providing statistical analysis of errors and failures.
* Establish a method for distributing reports to user groups, support groups, and managers as needed.

17.3 Organizing and staffing the storage administration group

Analyze who is currently performing storage administration tasks. Responsibilities for performing those tasks might be scattered throughout your enterprise. Systems developers, database administrators, operations and production control analysts, and users might be performing such tasks.

Consider the technical background and professional goals of your people when assigning responsibilities. By establishing educational requirements for your group, you build new skills and enhance existing one.

17.3.1 Storage administrators and storage administration groups

Storage administrators provide a major interface between differing groups in the data center. These groups have different requirements and agendas, and

it often falls to the storage administrator to produce a solution that satisfies all concerned.

17.3.1.1 Interaction with other groups
The following lists the major groups that a storage administrator could be expected to interact with.

- **End users** — They are customers of storage administrators and require services from them. Users often do not have the ability to articulate their requirements in a technical manner.

- **Application owners** — They determine the service required for data that they own and request services from the storage group to meet their requirements. They frequently request specific technical solutions.

- **System developers** — They provide services to the storage administration group by installing, maintaining, and modifying software used by the group. They create the hardware and software definitions for devices. System developers also frequently provide problem determination services for storage administrators.

  Systems developers are also customers of storage administrators although they often maintain responsibility for some storage functions on any volumes that they own. Systems developers may require specialized services of storage administrators for system data.

- **Operators/production support personnel** — They are users of storage administration services and may also be responsible for performing problem determination and implementing problem bypasses for the storage group.

- **Network administrators** — They enable the contact between the different organizations, through the network. The network administrators provide service to the storage administrator group and to the users. Distributed storage solutions such as Tivoli Storage Manager especially need particular service-level agreements with network administrators. The network administrators can, for example, decide to implement the Tivoli Storage Manager server on a faster network, or give advice for schedule times for heavy network-consuming processes.

  They are a key element of the distributed environments. As they use network analysis tools, they are a good point of contact, especially for performance troubles and for scheduling heavy network-use processes.

- **Database administrators** — They require services from the storage group to provide mechanisms for database placement, database performance, and backup. The storage administration group can help the database administrators to centrally manage their data by implementing

some functions, such as Tivoli Data Protections. These functions allow the centralizing management of online and offline databases.

- **Security administrators** — They give input to the storage administration group for security, like data encryption, physical access, and security categories.

- **Chief Information Officer (CIO) and Information Systems (IS) managers** — They manage all the information processing groups, and therefore lead in aiming for the goals and the way to reach them.

17.3.1.2 Skills required for storage administrators

This section focuses on skills specially required for storage administrators:

- General professional skills, such as leadership, project management, documentation, communication.

- Storage subsystems skills, like automated tape library (ATL), to understand technologies, standards, building blocks, and storage devices.

- Storage management software skills, to understand the functions, capabilities, and trends of your storage management software.

- Operating systems skills, to understand the user requests and manage the data stored onto these operating systems.

- Database management systems (DBMS) skills, to understand requests from database administrators and manage the data structure from these databases.

- Application and production skills.

- Other systems management software skills, such as scheduling, and network management software, which help automate and integrate storage administration task

17.3.2 Roles and responsibilities

In the following section, we discuss the roles and responsibilities that must be performed by any member of the storage administration group.

The only potential differences between a group with an enterprise scope and a user group managing a single disk pool are the additional complexities of managing data spread across multiple platforms and possibly multiple locations.

Although six different skill sets have been listed here, not all enterprises will be able to dedicate this number of people to storage administration. The

appropriate size for a storage administration group is discussed in topic 17.3.5, "Group sizing" on page 344.

Whatever the size of your enterprise, however, all the tasks identified will need to be performed by someone at some time. There is certainly scope to allow some tasks to be moved between differing levels, according to the skill set of the people who form your storage administration group:

- **Manager of the storage administration group** — As the manager of the storage administration group, this person works with your team to:

  - Justify establishing the storage administration group to management
  - Create a storage management mission statement
  - Develop short- and long-term objectives
  - Identify skills and education levels for the storage administration group
  - Define roles, responsibilities, and tasks for group members and leaders

- **Team leaders** — Team leaders act as a focal point within the storage group and should provide the major source of motivation for the entire group. Their responsibilities include:

  - Contact point for other groups
  - Resolving issues and concerns within the group
  - Reporting status of group work to management
  - Implementing the naming convention for Tivoli Storage Manager
  - Standardizing the descriptor used for archived data sets

- **Storage strategist** — The storage strategist is responsible for establishing project objectives and strategy planning. He/she is dealing the long term rather than in the short term. The storage strategist's responsibilities would include:

  - Evaluation of hardware and software, deciding which technologies may justify a practical review.

  - Evaluation of new versions and releases of the current software.

  - Verify the viability of their projects with the other groups.

  - Standards evaluation and implementation.

  - Risk assessment and documentation.

  - Recommend/justify hardware and software acquisition.

  - Updating the requirements plan to include new requirements as they become known. This can include collecting service requirements and maintaining the service-level.

- **Storage administrators** — The storage administration is becoming increasingly technical, as its association with the operating system

environment grows closer. It is essential that experienced systems technicians be a part of your group. Storage administrator's responsibilities include:

- Technical environment and service enhancements

- Translating service-level agreements into storage management policy definitions

- Maintaining and automating storage administration and management procedures based on development tools, for users and storage operator interfaces

- Problem control and resolution

- Second-level Help Desk service

- Service statistics for management reporting

- Monitoring performance with user, network, and hardware groups

- Monitoring the configuration of system

- Security and recovery

- Short-term service planning

- Perform capacity planning

- Design data recovery plans

- Practical evaluation of new hardware and software

- Commission new hardware and software

- Provide education for users and operations personnel

- **Operators** — The operators perform more routine storage management tasks. When they have gained more experience, they can assume some of the duties of storage administrators. The storage operator's responsibilities include:

  - First-level Help Desk support

  - Measurement and reporting

  - Performing daily routine storage management tasks for the storage subsystems, such as add or remove tapes from the libraries

  - Providing data recovery services for users and developers

  - Acting as liaison for user groups

  - Maintaining service-level agreements

  - Providing input to problem and change control

- Reporting on storage usage

- Monitoring storage management processes. For example, with Tivoli Storage Manager, the operators may monitor:

  • The scheduling of backups
  • The occupancy of the database and the recovery log
  • The availability of scratch volumes in the tape libraries
  • The recycling of the tapes onsite and offsite
  • The processes for database backup, and expiration, and updating the copy storage pools

• **Project managers** — The project managers may also be responsible for holding a watching brief over the groups input to projects being run by other departments. Their responsibilities would include:

- Verifying that the deliverable conform with the expectations

- Ensuring that all projects are delivered on time and within budget

- Detailed project planning

- Focal point of contact regarding groups involvement in projects

- Reporting project status to management

Project management may be a good intermediate step for a person with a technical background, who wishes to move into management. The project manager is a temporary job, which exists just during the implementation phase.

17.3.3 Assigning tasks

Consider that you have to structure the task assignments for your storage administration group by system-wide or installation-wide tasks. System-wide refers to a processor system, which is a system consisting of one or more processors and their associated disks, storage controls, and other devices.

In this structure, a single storage administrator assumes full responsibility for a single processor and all its associated storage subsystems. That person performs a wide range of tasks, thus developing a complete set of skills and experience. If you apply this to a large installation, be careful to ensure consistency across all systems.

Although you might assign most storage management tasks by system, it might not be feasible to manage the storage subsystem entirely on this basis. Some tasks might require agreement for the entire installation, such as maintaining policies and procedures. Also, you might call on other departments to provide assistance with specialized tasks on an

installation-wide basis. For example, the system programming group could assume responsibility for performance tuning.

17.3.3.1 Systems specialization — vertical skill backup

Plans for system-wide storage administration must include arrangements for members to act as backup storage administrators for other systems in the multisystem environment. Having a backup ensures that you consistently meet your service commitments.

You can accomplish this by a combination of the following approaches:

- Rotating storage administrators through the various system-wide and installation-wide tasks, thus cross-training them to cover any of the job assignments in your group
- Establishing backup responsibilities for each of your major job assignments
- Establishing consistent policies, naming conventions, and procedures throughout the installation. Installation-wide job assignments can enforce this consistency
- Clearly documenting job responsibilities.

17.3.3.2 Task specialization — horizontal skill backup

Another method of structuring storage administrator job assignments is to extend installation-wide responsibilities to include most storage management tasks.

In this framework, individual storage administrators specialize in a specific set of tasks across a complete installation, which can consist of one or more processors located in one or more system complexes.

For example, one person could be responsible for data set and object availability, handling tasks related to backup and recovery of data sets and objects. Others might handle space management and perform tasks related to data set and object deletion, data set conversion, and space usage.

These special assignments are by no means the only installation-wide assignments. You can, for example, establish a full-time position to handle user service requests and contacts with user group representatives.

A combination of installation-wide and system-wide job assignments can give you considerable flexibility. Increasing the number of installation-wide assignments lets you reduce the responsibilities of the storage administrators

who handle entire systems. Conversely, you could assign most tasks on a system basis and reduce installation-wide job assignments.

Factors that help you determine the job assignments include:

- Your organizational structure
- The skills of your people
- The extent of your personnel resources
- The degree of centralization of storage resources

17.3.4 Staffing

Analyze who is currently performing storage administration tasks. Responsibilities for performing those tasks might be scattered throughout your company. System developers, database administrators, operations and production control analysts, and users may be performing such tasks. Consider the technical background and professional goals of people when assigning responsibilities. By establishing educational requirements for your group, you build new skills and enhance existing ones.

17.3.5 Group sizing

The size and structure of your storage administration group depend on many factors:

- The storage capacity of your installation

- The number of differing platforms that need to be managed and the amount of automation available on each platform

- The number and types of applications

- The level of user knowledge

- The effectiveness of planning for the enterprise

- How much and how often your systems and users change

- Whether there are unusual or chronic problems, for example a poorly designed application that requires much manual intervention from administrators

- The management style of your company

The relationship between the numbers of storage administrators required and the amount of storage to be managed is nonlinear.

The most effective planning method is to meet with your group to discuss the expertise and intentions of each member and the needs of your installation.

You can then design your structure to meet business needs and, at the same time, distribute the responsibility appropriately among group members.

Although the following structures and titles are only suggestions, they can serve as a starting point. As your tasks become more defined and established within the installation, make the appropriate structural changes.

17.3.5.1 Structuring a small storage administration group

If your installation is small, your group could have one to three members and a DFSMS or Tivoli Storage Manager project implementation leader. As shown by Figure 112, the storage administration group should be placed within an existing department, such as technical support. The project leader reports directly to the department manager.
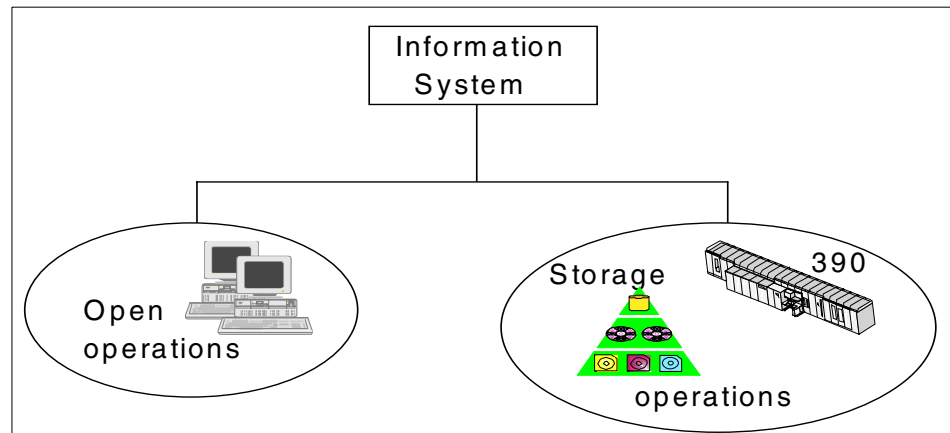
Figure 112. Localizing a small storage administration group

17.3.5.2 Structuring a medium-sized storage administration group

If your installation is of medium size, consider having a storage administration group of three to six members with one leader who is responsible for future implementation projects and service-related tasks. See Figure 113.
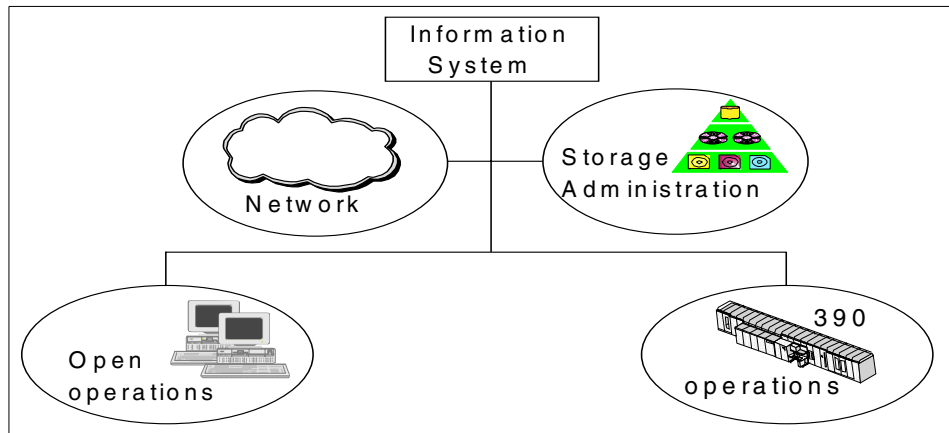
Figure 113. Localizing a medium storage administration group

17.3.5.3 Structuring a larger storage administration group

If you have a large installation, you should have a comparable number of storage administrators, between six and ten. As the group size increases, its structure becomes more complex, as shown in Figure 114.
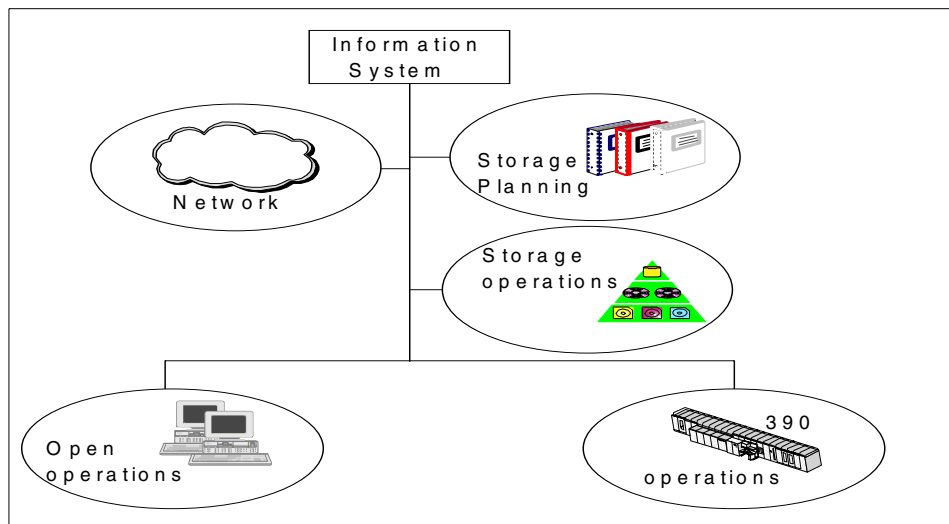
Figure 114. Localizing a large administration group

In a large installation, you can have two or more subgroups:

- Planning and support group, responsible for the technical implementation and support tasks as well as functions such as problem and change management

- Service delivery group, responsible for service-oriented tasks, such as capacity planning, security and disaster recovery, and storage administration services.

17.3.6 Group location

There are two separate factors to be considered when deciding where to locate storage administrators. The first is the physical location of the members of the group, and the second is the logical linkages of the group to other groups within the enterprise.

17.3.6.1 Physical location

The location of your storage administration group depends on the nature of your company. It is often most effective to have one concentrated group for the entire company. However, if you have an attended, remote site, you could have a storage administrator who manages data on that site. If the remote site is not attended, the storage administration group at the main, or parent, site handles storage administration tasks. Or, if your company is small, you might include storage administration tasks within existing departments.

17.3.6.2 Logical location

The logical linkage of the storage administration group will largely be determined by the number of staff that will report to the group. A small group may not be considered large enough to exist on its own, and the members might report to a technical group, or an operations manager.

It is important that the group is not seen as tied to a particular operating environment. Reporting through the technical structure is recommended over reporting through the operations pathway, because of the increasing skill requirements of the group.

If the group is too large for a single manager, then the lines of demarcation into smaller groups can be influenced by either the types of work being performed, or the range of systems being managed; for example, you could create an availability group managing backup processing, or you could have one group supporting finance systems and another supporting order entry. The decisions that you make will also be influenced by the skill sets that you have identified for the storage group.

17.3.6.3 Education requirements

Your storage administrators must understand the capabilities of the hardware and software components that they are using daily.

Chapter 18. Exploiting storage solutions services

IBM Global Services (IGS) is real people solving real problems, strategists, implementors, people who help you make sense of technology — people who will make sure that the business solution you want is the solution you get. An extension to this world-wide group is our Business Partner network, which enables us to address all our customers even more efficiently.

Figure 115 shows the IGS home page. The URL is as follows:

```
http://www.ibm.com/services
```


Figure 115. IBM Global Services homepage

18.1 Professional Services

IBM Global Services offers a wide range of services to support both hardware and software products at every stage of the product life cycle.

18.1.1 Technical Support Services

The following Technical Support Services are available.

18.1.1.1 Installation Services

- Installation Services for Geographically Dispersed Parallel Sysplex for Extended Remote Copy (GDPS/XRC)

  To help you address the need for availability, IBM Installation Services for GDPS/XRC combines technical expertise with automation technology. From working with you to plan availability requirements to implementation and testing, we offer a wide range of services to help you improve disaster recovery capabilities.

  Highlights:

  - Protects your data by helping you plan and install remote data copying for S/390 servers
  - Allows you to reduce costs and speed implementation with upfront planning
  - Provides solutions for your unique environment by supporting a range of transaction and database managers
  - Offers flexible service options to enhance application and data availability

- Installation Services - Installs IBM 3583 Ultrium Scalable Tape Library on Host Server or Storage Area Network

  The IBM 3583 Ultrium Scalable Tape Library will enhance your SAN environment whether new, existing, or planned. This tape library is designed for installation by customers but because of the complexities of implementing or expanding a SAN environment or resource limitations, you may want IBM assistance with the installation of your 3583 Ultrium tape library. IBM provides physical planning and installation assistance to help you get your device up and running quickly.

  Highlights:

  - Delivers a speedy, smooth installation
  - Gets your tape drive into production sooner
  - Helps keep your staff focused on business-critical activities

- Installation Services for SANergy

  IBM Installation Services for SANergy enable file sharing among multiple SAN disk storage devices. By sharing storage resources within a disk subsystem or across multiple IBM and non-IBM subsystems, you can enhance network efficiency by assigning storage capacity to independent file systems. Resources can be shared across heterogeneous UNIX, Microsoft Windows NT and Microsoft Windows 2000 operating systems.

Highlights:

- Improves access to centralized databases that support mission-critical applications

- Improves flexibility by enabling cost-effective network performance

- Implements SAN solutions within your LAN environment to improve network efficiency

18.1.1.2 Hardware Maintenance Services

IBM Global Services offers a range of maintenance services for both IBM and non-IBM hardware, providing a single source you can depend on for all your hardware support needs. Our worldwide, state-of-the-art service delivery system brings it all together to ensure your hardware is up and running, fast.

Highlights:

- 24x7 support

- Knowledgeable remote and onsite service specialists

- Experienced service planners

- Global databases with vast machine service and problem resolution histories

- Worlds largest parts inventory

- Advanced remote diagnostic and communications tools

- Predictive and preventive maintenance

18.1.1.3 Migration Services

The following Migration Services are available:

• Migration Services for S/390 attached IBM storage servers

Migrating data can consume your staff 's time and interrupt systems operations, resulting in costly downtime or data loss. We can help reduce these exposures by having one of our services specialists perform the migration, allowing your staff to concentrate on business-critical activities.

We migrate data from 3990-compatible products to an IBM storage server with minimal disruption to your work environment.

Highlights:

- Minimizes costly downtime and data loss

- Allows movement to and from 3990-compatible storage systems without special configuration changes

- Provides freedom to combine or split subsystems or move portions of data in an existing storage subsystem
- Superior data protection preserves data updates throughout the migration, allowing the process to be interrupted if needed

• Migration Services for Virtual Tape Server for New Allocations

Now you can effectively manage your expanding tape library. Use IBM Migration Services for Virtual Tape Server for New Allocations to control your tape library. Each time you execute your backup strategy, our virtual tape server (VTS) storage software creates a new storage allocation. Let our knowledgeable and experienced storage specialists help you set up your tape volume stacking requirements.

Our storage specialist has the experience to change the way you archive data on your automated tape library storage system. Let us use our experience and storage skills to migrate your applications.

Stacking data volumes on a single tape cartridge reduces your tape library space requirements. By employing this service to reengineer your new tape allocations, your future tape storage will be more efficient. Our storage specialist teaches your staff how to complete your VTS solution.

Highlights:

- Helps improve your tape library management techniques
- Gives you access to the knowledge and experience of a qualified IBM storage specialist
- Helps you accomplish your core business needs while learning how to complete your virtual tape server implementation
- Offers cost savings opportunities to your organization

• Migration Services for Virtual Tape Server for Static Data

Many enterprises have the same issues with their tape libraries: tape storage takes up valuable floor space, physical space on your tape is often wasted, and your staff may be using a tape management methodology that no longer suits your needs.

Use IBM Migration Services for Virtual Tape Server for Static Data to help control your tape library space requirements and improve your tape volume stacking. Let our storage specialist implement and lead your effort to reduce library space and to train your operators on the handling of existing tapes.

Highlights:

- Moves your existing tape volumes to your virtual tape server (VTS)

- Reduces your tape library space requirements

- Offers you access to an experienced IBM storage specialist

- Helps you accomplish core business needs while learning how to complete your VTS implementation

- Offers cost savings opportunities to your organization

18.1.1.4 Operational Support Services

Our expert services specialists will show you new and expanded ways to get more from your storage technologies. By analyzing your storage resources, IBM experts can determine how your applications can benefit from this special optimization service. Whether you want to explore alternate methods or learn the latest techniques, our skilled specialists can assist you in achieving your goals.

Highlights:

• Helps you optimize use of your existing storage resources

• Helps you maximize your investment in current and future storage technology

• Provides you with knowledge of new features and functions

• Helps you achieve greater storage efficiencies and cost savings

The following list shows operational support services available for each storage solutions:

• Operational Support Services for 3494 Automated Tape Library with Virtual Tape Server

• Operational Support Services for ADSM for SAP

• Operational Support Services for RAMAC Virtual Array and Snapshot

• Operational Support Services for Tivoli Storage Manager

• Operational Support Services for Tivoli Storage Manager for DRM

• Operational Support Services for Tivoli Storage Manager for Network Storage Manager

• Performance Management and Capacity Planning Services for Tivoli ADSM

18.1.1.5 Power Protection Services

IBM Global Services offers power protection solutions to meet your needs for uninterruptible power supply (UPS) and surge management protection. Our

power protection specialists' consultation services include planning, sizing, installation coordination, software monitoring, and testing.

18.1.1.6 Site Services

- Data Center Services

Data Center Services can help maximize your information technology investment by reducing complexity, simplifying vendor management, and streamlining your operations so you can redirect your focus on your core business objectives. Improve your availability, reliability, performance, and end-user satisfaction through IBM's portfolio of services and technical expertise.

Highlights:

- Reduce complexity
- Simplify vendor management
- Streamline operations
- Improve skills
- Redirect focus to core business objectives

- Data Center Planning, Design and Construction Services

In today's rapidly changing information processing environment, it is important to plan your hardware configuration requirements and future growth to keep up with the current technology. Poor planning can result in increased costs, timely delays, needless changes, annoying disruptions, and unnecessary operational downtime. When making physical changes to your information processing environment and facility, controlling costs, managing budgets, and scheduling and meeting deadlines become critical success factors... factors that can very well make the difference between success or failure! Do you have the time and the resources to tackle a project so complex?

Highlights:

- Offers everything from preliminary consulting to a turnkey solution
- Helps you manage change in your physical environment
- Multivendor services provide you with total support
- Offers world-class quality in a cost-effective, timely solution
- Analyzes your current computing environment and project growth

- Site Services

IBM Site Services offers a wide range of services to help make your transition to the e-business market space a success. From initial assessment to design and installation services, we can help restructure your physical IT environment for improved performance. Whether you are

planning a new IT facility, retrofitting an existing IT facility or workspace, or preparing your facility for security or intelligent building systems, we can help you with a solution.

Highlights:

- Helps reduce risk and improve availability by assessing your environment and providing recommendations

- Assists in preparing your site for future needs and helps reduce planning costs with flexible solutions

- Allows you to concentrate on keeping your business running smoothly by leveraging experience

- Helps ensure that your project is completed on time and on schedule with tailored solutions

18.1.2 Networking and Connectivity Services

18.1.2.1 Connectivity Services

The following Connectivity Services are available:

- Fiber Transport Services

Are you overwhelmed with the maze of unwieldy and cluttered underfloor wiring? Are you in the process of migrating from a parallel computing environment to the serial optical environment of ESCON? Think IBM FTS! We have designed a system especially for the optically-connected data center.

Highlights:

- Structure your channel cabling system to maximize ease of future reconfigurations, relocations, and additional hardware

- Eliminate unwieldy point-to-point underfloor cluttered wiring

- Leverage your Enterprise System Connection Architecture (ESCON)

- Reduce risk... simple, flexible, and economical migration and installation

- Use multifiber terminated push-on MTP technology for fast installation

- Reduce configuration costs

- Your satisfaction is guaranteed!

18.1.3 Testing Services

The following Testing Services are available:

18.1.3.1 Testing Services

- Testing Services for SAN

    See 18.3, "SAN Interoperability Lab" on page 360 for more details.

18.2 Learning Services

Learning Services is the other main domain of activity of the IBM Global Services Organization. As for Professional services, refer to 18.1, "Professional Services" on page 349 for more information. The scope of Learning Services is very wide, addressing education for IBM as well as non-IBM products and solutions. The Learning Services portfolio can be accessed through the URL shown below:

```
http://www.ibm.com/services/learning
```

From the screen shown in Figure 115, you should select IBM Learning Services, and then for local organization, select your country. Figure 116 shows the screen you get when choosing the IBM Learning Services option. International events and conferences are also advertised in this URL.
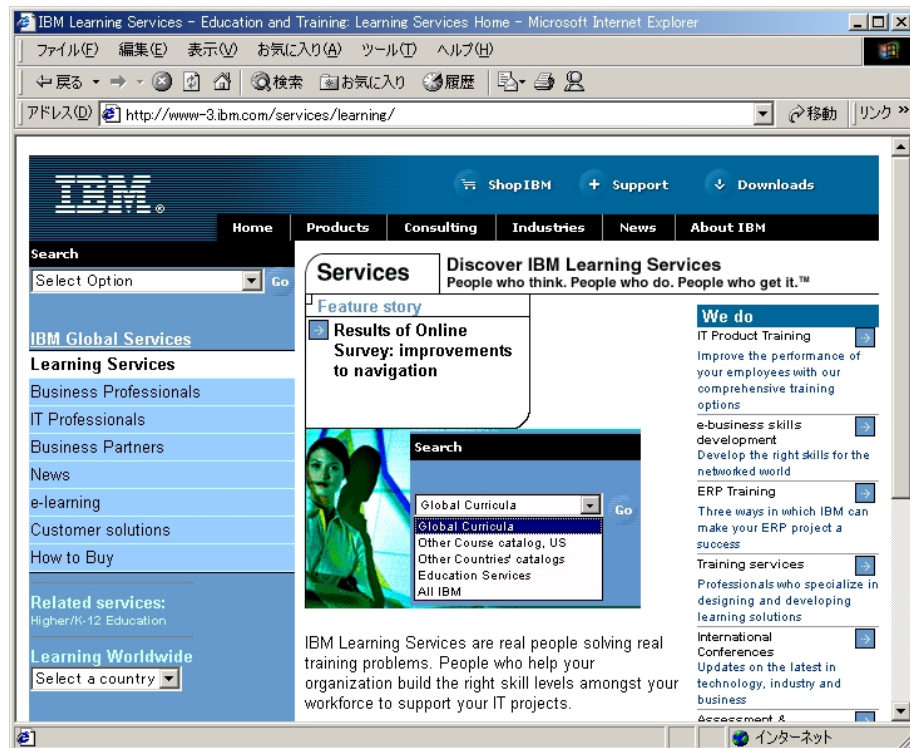
Figure 116. The Internet screen for IBM Learning Services

International workshops around the world are also taught by ITSO experts (such as those who published the redbook that you are currently reading) — bringing you the latest technical information about all our products. These ITSO workshops are also advertised through this Internet site.

Learning Services can be performed at IBM locations as well as at customer locations, where specific classes can be developed and taught to address your exact expectations, and to minimize your travel expenses. Here is a list of some classes and class numbers related to the storage arena:

- Introduction to Data Storage (J3700)
- Storage Management Fundamentals (J3701)
- DFSMS Implementation (J3711)
- DFSMShsm Implementation (J3717)
- DFSMShsm: ABARS Workshop (J3719)
- Managing DFSMShsm Recovery (J3737)
- DFSMS Optimizer Implementation (ES67A)

As for Professional Services, our Business Partner network is a natural extension of our Learning Services forces, making our offering even more interesting to match or exceed your learning objectives.

To better understand the richness of our Learning Services offering, here are two new detailed offerings that should interest your organization:

- 18.2.1, "Distributed learning" on page 358
- 18.2.2, "Professional certification program" on page 359

18.2.1 Distributed learning

Here is the dilemma: you want to make the move to a network-based business. And that means training your people. But how can you quickly educate your people on a new way of doing things when you are still using traditional training?

IBM Learning Services can show you how to move your people and your business to a network-based system using the latest distributed learning technologies in the following sections:

- 18.2.1.1, "Benefits of distributed learning" on page 358
- 18.2.1.2, "When you would use distributed learning" on page 359
- 18.2.1.3, "Examples of IBM distributed learning technology" on page 359

18.2.1.1 Benefits of distributed learning

Distributed learning solutions play an essential role in today's education and training challenges by:

- Cost savings:

    - Train faster
    - Reduce/eliminate travel costs
    - Decrease per-student training expense

- Greater impact:

    - Just-in time-learning
    - Learning when you work
    - Extended access through networks

- Higher quality:

    - Interactive, customized
    - Focus and flexible
    - More effective, more easily retained learning

18.2.1.2 When you would use distributed learning

Typical scenarios when distributed learning is appropriate are:

- Decrease the total cost of training

- Need to train large numbers of people

- Work force is decentralized or mobile

- Training environment is dynamic and volatile, in line with rapidly changing business needs

- Must train quickly or be outpaced by competitors

- Need flexibility (requiring customized solutions).

18.2.1.3 Examples of IBM distributed learning technology

- Self-teach — for example, CD-ROMs that students can work through anytime, any place

- Instructor-led — for example, self-teach, but with access to an instructor and with interaction with other course participants, even in different locations worldwide

- Briefings — for example "live" broadcasts of presentations, with options for student feedback and questions

- Performance support — detailed help available on demand, for example, to help you understand functions of a new application

- Administration — monitoring and tracking students' progress through courses electronically, and self-assessment tools.

18.2.2 Professional certification program

Today's ever-changing business environment, it's essential to keep pace with complex technologies and products. Most businesses are looking for complete solutions rather than tactical efforts. The Professional Certification Program from IBM offers a business solution for skilled technical professionals seeking to demonstrate their expertise to the world.

The program is designed to validate your skills and demonstrate your proficiency in the latest IBM technology and solutions. In addition, professional certification may help you excel at your job by giving you and your employer confidence that your skills have been tested.

The Professional Certification Program from IBM provides you with a structured program leading to an internationally recognized qualification. The program is designed for flexibility by allowing you to select your role, prepare for and take tests at your own pace, and, in some cases, select from a choice of elective tests best suited to your abilities and needs. Some roles also offer a shortcut by giving credit for a certification obtained in other industry certification programs.

The following list shows the individual certification programs you can choose from the Storage Systems and Storage Management area.

- IBM Certified Specialist - Storage Area Network (SAN) Foundations
- IBM Certified Specialist - Storage Area Network (SAN) Implementation
- IBM Certified Specialist - Enterprise Tape Solutions
- IBM Certified Specialist - Enterprise Disk Solutions
- IBM Certified Specialist - IBM Storage Sales
- IBM Certified Specialist - IBM Mid-Range Storage Technical Specialist

18.3 SAN Interoperability Lab

Figure 117 shows the SAN Interoperability Lab.

Figure 117. SAN Interoperability Lab

Appendix A. Some previous disk storage subsystems

This appendix addresses some previous disk subsystems that are referenced in the book, and that are replaced or are about to be replaced by new subsystems.

A.1 Fiber Channel Storage Server

In this section we present a high-level overview of the IBM Fibre ChannelRAID Storage Subsystem (FC-RSS) and its supporting software. For more information, see *Configuring and Implementing the IBM Fiber Channel Raid Storage Server*, SG24-5414.

The FC-RSS consists of dual controllers, up to six drawers of disk drives, and optionally, one or two Fibre Channel hubs mounted in an IBM Seascape Solutions Rack (see Figure 118).

Figure 118. IBM Fibre Channel RAID Storage Server

A.1.1 Overview

The FC-RSS is an entry-level to mid-range solution for customers with a need for a multiserver storage subsystem that has a capacity of 36 gigabytes to one terabyte of RAID protected storage in a single rack. It provides mission-critical, easy-to-manage high availability storage with redundancy in the controllers, RAID disk storage, power supplies, and cooling units. Customers with storage needs that are supported across multiple homogeneous server platforms and that require Fibre Channel Arbitrated Loop (FC-AL) attachment should consider the FC-RSS.

The IBM Enterprise Storage Server (see 7.1, "The IBM Enterprise Storage Server (ESS)" on page 121) and the IBM 7133 Serial Disk Subsystems (see section 7.2, "7133 Serial Disk Subsystem Advanced Models D40 and T40" on page 132) should be considered for large enterprise-wide heterogeneous platforms, or for RS/6000 clustered environments where appropriate adapters may offer the right functions and storage with price performance characteristics that better meet customers' requirements. The FC-RSS provides users a Fibre Channel based storage subsystem that allows centralization of disk storage, systems management, and a facility for data sharing in a homogeneous system environment.

The FC-RSS offers a high-performance, high-availability storage solution for servers running Microsoft Windows NT, HP-UX, Sun Solaris, or AIX. Its industry standard Fibre Channel host attachment provides configuration flexibility for distances of up to 10 kilometers. It offers enhanced data availability through the implementation of various RAID techniques, disk drive module sparing, failover protection, and dual paths to data.

A.1.2 Architecture

Figure 119 shows the architecture of the FC-RSS in schematic form.
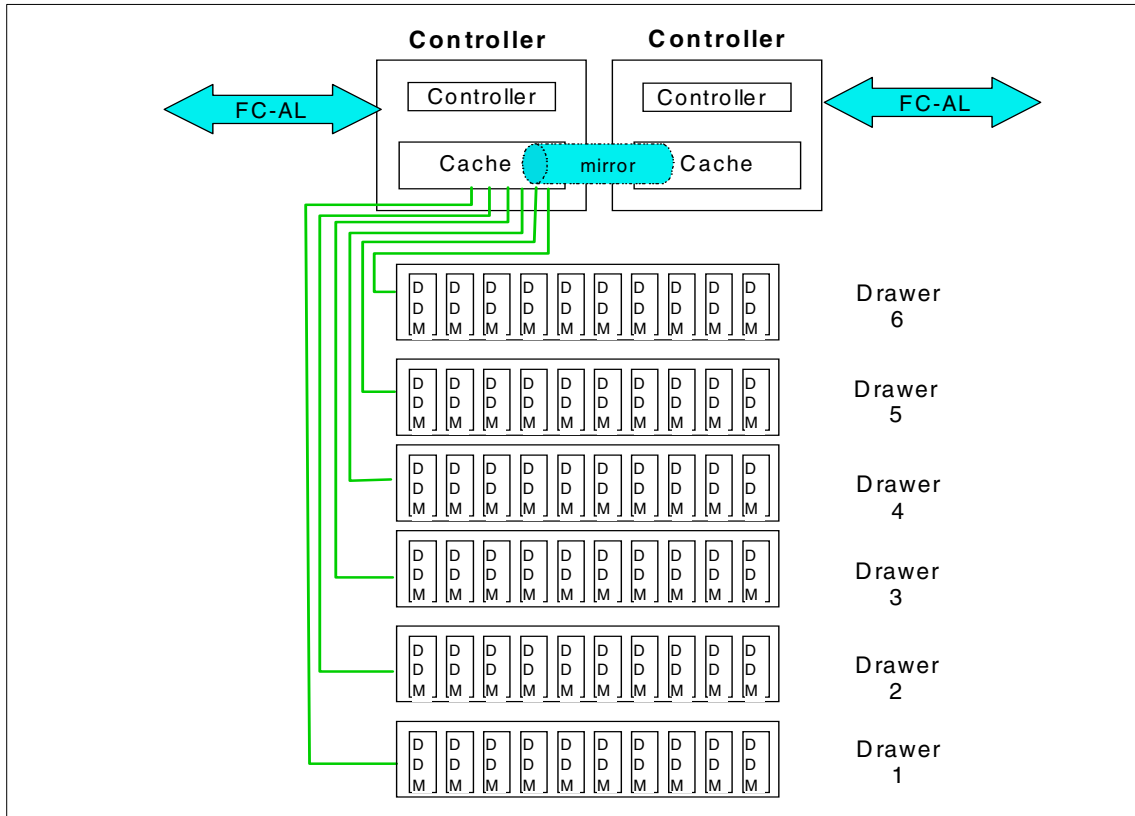
Figure 119. IBM Fibre Channel RAID Storage Server Architecture

A.1.3 Key features

The FC-RSS provides the following interesting features:

- Multihost attachment (see A.1.3.1)
- High-availability design (see A.1.3.2)
- Extended distances (see A.1.3.3)
- High performance (see A.1.3.4)
- Dynamic management (see A.1.3.5)
- Expandable storage unit (see A.1.3.6)

A.1.3.1 Multihost attachment

The IBM Fibre Channel RAID Storage Server is a robust storage solution for small clusters of like servers in environments requiring Fibre Channel-Arbitrated Loop (FC-AL) attachment. The IBM Fibre Channel RAID Storage Server can be shared by two homogeneous Windows NT or UNIX servers.The Fibre Channel RAID Storage Server is ideal for midrange servers running critical applications that require a storage system with built-in redundancy and fault-tolerance.

A.1.3.2 High-availability design

As access to information becomes more critical than ever, companies are demanding flexible, high-performance, continuously available storage systems. With RAID levels of 0, 1, 3, 5, and 0+1 to choose from, the IBM Fibre Channel RAID Storage Server can help you reach the optimum system performance and protection for your business-critical applications. Its dual-active RAID controllers provide high throughput and redundancy. Both controllers can be active at the same time and still provide failover in case of emergency, without any intervention. To increase availability, each RAID controller supports up to 256 MB of battery-backed cache. Dual fans and power supplies further contribute to its ability to support 24 x 7 operations. Components such as controllers, fans, drives, and power supplies are hot pluggable, simplifying access and maintenance for continuous operations.

A.1.3.3 Extended distances

Industry-standard Fibre Channel host attachment provides the flexibility to configure the Fibre Channel RAID Storage Server at distances of up to 10 km from the host server, which is suitable for disaster recovery implementations.

A.1.3.4 High performance

High-speed Fibre Channel attachment meets the requirements of high-performance networked systems running large database applications, such as enterprise resource planning and data warehousing. The IBM Fibre Channel RAID Storage Server has dual Fibre Channel ports, which give an aggregate bandwidth of 200 MB/sec. With dual FC-AL Windows NT hosts, sustained data transfer rates of up to 179 MB/sec can be achieved with 256 KB sequential reads. Over 16,200 I/O operations per second can be reached with 4 KB 100 percent cache hits.

A.1.3.5 Dynamic management

The StorWatch Fibre Channel RAID Specialist is a network-based integrated storage management tool that lets storage administrators configure, monitor, dynamically change, and manage multiple Fibre Channel RAID Storage Servers from a single Windows95 or Windows NT workstation (see Chapter 12. StorWatch on page 245). High availability and full redundancy are provided with the host-specific Fibre Channel Storage Manager tool, which resides on the host system and provides automatic I/O path failover when a host adapter, IBM Fibre Channel Storage Hub, or a storage server controller fails.

A.1.3.6 Expandable storage unit

The IBM Fibre Channel RAID Storage Server supports the new IBM Expandable Storage Unit, a disk drawer designed to meet your growing storage requirements. It features IBM high-performance Ultrastar* disk drives in capacities of 9.1 GB, and 18.2 GB. Storage capacity can start at 36 GB and grow to over 1 TB by placing a maximum of six Expandable Storage Units in a single IBM Seascape* Solutions Rack. The IBM Seascape Solutions Rack also houses the Fibre Channel RAID Storage Server and the IBM Fibre Channel Storage Hub(s), all in an industry-standard 19-inch rack.

A.1.4 Specifications

This section addresses:

- The IBM Fibre Channel storage solution (see A.1.4.1)
- The product characteristics (see A.1.4.2)
- Software components (see A.1.4.3)
- Supported platforms (see A.1.4.4)

A.1.4.1 IBM Fibre Channel storage solution

The components of the IBM Fibre Channel storage solution include the following:

- 2102-F10 IBM Fibre Channel RAID Storage Server — Dual active RAID controllers support up to 512 MB of battery-backed dynamic cache. RAID levels of 0, 1, 3, 5, and 0+1 are supported. Redundant power and cooling are standard.

- 2102-D00 IBM Expandable Storage Unit — Supports up to 10 IBM Ultrastar drives (Ultra2 SCSI LVD) per unit. Each unit requires a minimum of four drives with similar capacities. Up to six Expandable Storage Units can be attached to a Fibre Channel RAID Storage Server. Redundant power and cooling are standard.

- 2103-H07 IBM Fibre Channel Storage Hub — Provides flexibility when configuring multiple host or Fibre Channel RAID Storage Server attachments, and enables greater distances to the FC-AL topology. The standard four short-wave optical Gigabit Interface Converter (GBIC) ports enable fiber cable connections of up to 500 meters Optional long-wave GBIC ports enable configurations as far as 10 km. The hub can be rack-mounted or used in a tabletop setup.
- 2101-100 IBM Seascape Solutions Rack — Can house an IBM Fibre Channel RAID Storage Server, six Expandable Storage Units and two IBM Fibre Channel Storage Hubs. Optional dual-power distribution units and dual-line cords are available.

A.1.4.2 Product characteristics
Product characteristics include the following:

- Storage capacity per expandable unit (2102-D00):
    - 36.4 GB as a minimum
    - 182 GB as a maximum
- Storage capacity per FC-RSS (2102-F10):
    - 36.4 GB as a minimum
    - 1092 GB as a maximum
- Optional long-wave GBIC ports enable configuration as far as 10km
- Broad range data protection options with RAID 0, 1, 3, 5, 0+1
- High performance with a bandwidth of 200 MB/sec
- Storage management with StorWatch Fiber Channel RAID Specialist
- High availability through automatic I/O path failover
- Concurrent maintenance with hot-pluggable components:
    - Controllers
    - Fans
    - Drives
    - Power supplies
    - 512 MB of battery-backed dynamic cache.

A.1.4.3 Software components
The FC-RSS software provides for configuration setup and subsystem management.

The IBM Fibre Channel Storage Manager provides platform specific automatic failover management. FC-RSS systems administration and the IBM Storwatch Fibre Channel Raid Specialist allows you to perform systems administration functions as well as manage multiple FC-RSSs from a single network attached workstation.

The dynamic FC-RSS management software includes the following features:

- Logical unit (LUN) creation and deletion
- Dynamic capacity expansion
- RAID level migration
- Segment size changes
- Defragmentation
- Status and performance monitoring
- Programmable error notification
- Maintenance and tuning utilities
- Step-by-step recovery procedures

A.1.4.4 Supported platforms

The FC-RSS supports these IBM and non-IBM systems:

- Intel-based servers running Windows NT Server 4.0
- Sun Microsystems SPARCserver and Enterprise servers with Sun Solaris
- 2.5.1 and 2.6
- Hewlett Packard 9000 K Series server with HP-UX V10.20, 11.0
- RS/6000 with AIX 4.3.2

Details on models and operating system releases are available on the Web at this URL:

```
http://www.ibm.com/storage/fcss
```

A.2 Versatile Storage Server

Versatile Storage Server (VSS) is designed for customers who require a central pool of storage. It is the first disk product that uses the Seascape architecture. It supports the concurrent attachment of IBM and non-IBM servers.

A.2.1 Overview

Versatile Storage Server, with the following features, consists of a high-performance storage server (shown in Figure 120):

- The storage server consists of two clusters with RISC symmetric multiprocessing (SMP) configurations, the host adapter, PCI Link technology, SSA adapter and communication ports.

- Cache is 512 MB, expandable to 6 GB

- One or more drawers of high-performance SSA disks and adapters using RAID-5 technology. The minimum usable storage capacity is 230 GB, with two drawers (2 drawers * 13 drives * 8.77 GB per drive). The maximum usable storage capacity is 2 TB, with 18 drawers (18 * 13 * 8.77 GB per drive).

- Redundant power supplies from S/390 servers.
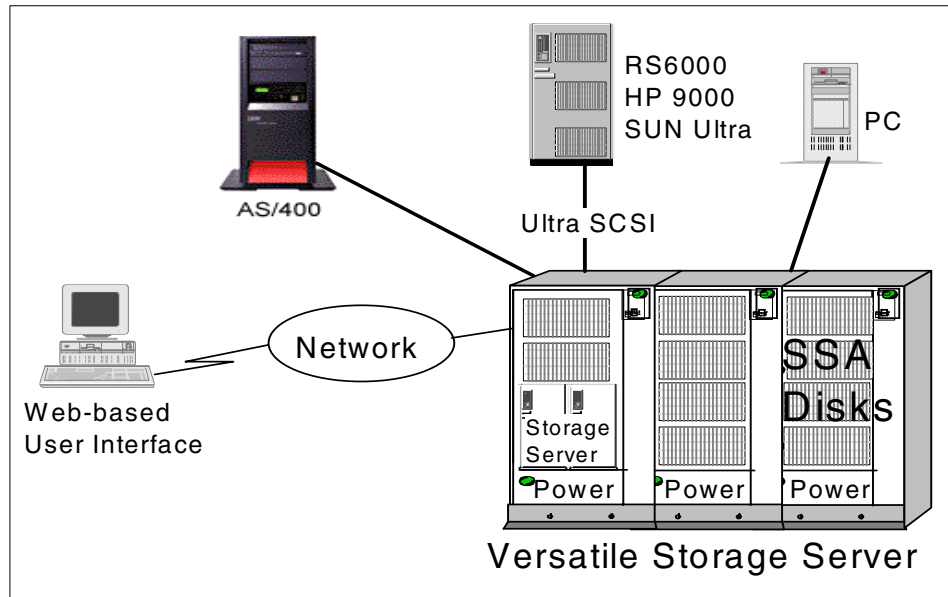
Figure 120. Versatile Storage Server architecture

VSS provides a Web-based interface, known as the IBM StorWatch Versatile Storage Specialist, for user configuration and control of the subsystem, and a serial interface for use by service personnel. Additionally, a modem connection can be made for remote diagnostics and download of fixes.

VSS can be configured as a single-rack or multiple-rack complex. The 2105-B09 storage server is always the first rack and includes the storage server, disk drawers, and power control system. The 2105-100 expansion enclosure includes space for 7133 drawers/disks and a power control system.

It is possible to combine the primary enclosure with up to two expansion enclosures.

A.2.2 Key features

VSS offers the customer the following key features:

- A central pool of storage distributed among up to 64 heterogeneous hosts
- The ability to assign a storage partition to multiple hosts so that true on-line data sharing can take place
- Built in protection against data loss, and loss of data integrity
- An easy-to-use Web-based interface, so that configuring and maintaining the system is as simple as possible. The interface is called Versatile Storage Specialist (VS Specialist).

A.2.3 Specifications

Supported hosts are the RS/6000, AS/400, Sun Ultra series, the HP/9000 800 series, some Data General models and also PC support on various IBM and Compaq servers.

Host attachment is by the use of UltraSCSI adapters; this is backward compatible with SCSI-2D F/W. The VSS presents a known disk image to the attached host. The AS/400 sees the VSS with the image of a 9337 disk subsystem while UNIX machines see a generic SCSI disk. The disk image can be divided into logical units (LUN).

A.2.3.1 Centralized storage

The storage can be easily expanded or reassigned among the attached servers as the needs of the business change. Existing 7133 SSA disk drawers can be attached to VSS, protecting the customer's existing storage investment.

A.2.3.2 Data sharing

Each server can have a dedicated portion of the storage, or the storage can be shared among the UNIX- or Windows NT-based homogeneous servers, using the appropriate application code or database manager to regulate concurrent access (read or write). File locking and record-level locking are the responsibility of the application program.

For example, a 16 GB logical volume can be assigned to two or more UNIX-based servers running the same OS. If both servers are using software that regulates access to the data (that is, Lock Management), then data could be written from one server and queried from the second.

A.2.3.3 Data protection
Data protection is provided through the use of RAID-5 technology and the dual-loop SSA disk adapter card. A RAID-5 array consists of eight disk drives, up to two arrays per drawer and consists of one of the following configurations: six-drives-plus-parity-plus-spare (6+P+S) or seven-drives-plus-parity (7+P). Each RAID-5 array must use the same capacity drives, either 4.5 GB drives or 9.1 GB drives.

Through its plug-and-play I/O architecture, components can easily be changed nondisruptively as technology changes. Hot spares are available across the loop. Should a disk drive fail, the rebuild process is started immediately, using the spare drive. The hot spare can float across the loop, thus eliminating the need to copy data back to the replacement drive after the repair is completed.

A.2.3.4 Failover protection
Should a processing cluster fail, all transactions are routed to the second cluster (failover). With cluster failover, RAID-5 protection, and redundant power, the servers provide high-availability operations for most types of failures.

Automatic use of a secondary path can be provided in case of adapter failure. The VSS needs connections to two different host interface adapters. For greatest availability, host-initiated software mirroring (with the exception of AS/400) should be implemented with dual paths configured from each host server to the VSS complex.

A second alternative to provide automatic use of a secondary path is the use of HACMP (for RS/6000), ServiceGuard (for HP) or similar software for non-IBM servers should a failure occur.

Most failures can be repaired without disrupting data access. A failed disk adapter, host adapter, or cluster can be replaced or repaired without shutting down the entire subsystem. While certain failures may disrupt access to a portion of the data, access to the remainder of the data is not affected.

A.2.3.5 Read/write cache

VSS includes read/write cache and fast write cache to enhance the system performance.

If a sequential access pattern is detected, intelligent algorithms monitor the type of data access, begin to fetch additional data before it is needed (prefetch) and place it in the processor cache. When the next request is processed, the data is immediately sent to the server, thus providing optimum performance.

When data is sent from the server, it is initially stored in the storage server cache and immediately transferred to the nonvolatile fast write cache. Once data has been placed in the nonvolatile fast write cache, the server is notified that the transaction is complete. Once the data is safely stored, it will be scheduled for destaging to the SSA disk drives.

The amount of read/write cache that is required depends on the applications (sequential applications are more cache friendly than random access applications), the amount of server memory available for the application, and the memory management philosophy of the server operating system.

A.2.3.6 Power subsystem

Each VSS rack includes a power subsystem with redundant components, that, in conjunction with the dual processing clusters and host interface adapters, make for a fully redundant power subsystem. No single point of failure within the power subsystem should affect availability of VSS. The power subsystem includes redundant line cord input, main power distribution, and cooling fans.

A.2.3.7 IBM StorWatch Versatile Storage Specialist

Using a HyperText Transmission Protocol browser (http), you can access the VS Specialist. It provides the ability to view and update the configuration and monitor error logs, and to set up and change user profiles and problem notification parameters (phone numbers, e-mail addresses, and SNMP routing).

A.2.3.8 IBM Versatile Storage Server Service Alert

VSS includes a set of features designed to provide rapid diagnostic and remote support capability. The customer is responsible for providing an analog phone line near the Versatile Storage Server to enable remote diagnostic support.

The VSS Service Alert includes the following functions:

- Provides continuous self-monitoring, which will initiate a call to the service provider if a failure happens or is about to happen.

- Allows the service provider to initiate a call to the VSS processor and remotely analyze potential or actual failures, then determine the type of failure and proper fix.

- Allows VSS to correct many types of microcode problems. When a problem is reported, a fix can be created and then applied from the remote location.

See *IBM Versatile Storage Server*, SG24-2221 for more information.

A.3 RAMAC 3 Array Storage

RAMAC 3 Array Storage offers the highest level of data availability through its third-generation RAID 5 technology and disaster protection functions.

RAMAC 3 Array Storage is illustrated in Figure 121.

Figure 121. RAMAC 3 array storage

A.3.1 Overview

RAMAC 3 array storage provides the highest levels of data availability.

The RAMAC 3 array storage subsystem consists of RAMAC storage control and one or two RAMAC 3 storage frames, each containing up to 363.2 GB of storage. A fully configured RAMAC 3 storage frame contains 16 RAMAC 3 drawers, each with 22.7 GB of storage capacity, giving a total of 363.2 GB. Therefore, two RAMAC 3 storage frames each containing 16 drawers provide 726.4 GB. The minimum configuration is a RAMAC 3 storage frame containing two RAMAC 3 drawers, providing a total capacity of 45.4 GB.

There are two models of the RAMAC3 storage control unit: RAMAC3 storage control Model 001 (9390-001), which contains one storage control, and RAMAC 3 storage control Model 002 (9390-002), which contains two storage controls. Each storage control, independently, connects to one of the storage frames.

The minimum cache size for a storage control is 256 MB. This size ensures that the storage subsystem delivers good performance for lower-capacity configurations. Maximum cache is 8 GB (4 GB per storage control in a 9390-002).

The nonvolatile storage (NVS) sizes vary from a minimum of 32 MB (on a 9390-001 single storage control) to 256 MB (128 MB per storage control on the 9390-002).

Each storage control on a 9390-002 supports up to 128 devices for a total of 256. Each of the two RAMAC 3 frames contain from 2 to 16 drawers, each emulating eight logical 3390-3 volumes.

9390 storage control supports only ESCON channel attachment. The minimum channel configuration is four ESCON channels on a 9390-001 and the maximum is 32 on a 9390-002 (16 on each storage control). The maximum simultaneous host data transfer is eight for 9390-002 (four for each storage control).

See *IBM RAMAC 3 Array Storage*, SG24-4835 for more details.

A.3.2 Key features

RAMAC 3 includes the following key features:

- A.3.2.1, "RAID 5" on page 374
- A.3.2.2, "Remote Copy" on page 374
- A.3.2.3, "Advanced cache management" on page 378
- A.3.2.4, "Concurrent Copy" on page 378
- A.3.2.5, "3990-6 investment protection" on page 381

A.3.2.1 RAID 5

The RAMAC 3 drawer is designed for outstanding data availability. The drawer offers third-generation *RAID 5* protection and the most reliable microcode for its disk drives. The other hardware components, including the microprocessors, are fully redundant and fault tolerant. All electronic components are fully backed up by battery, and the battery itself is checked by the control unit every 24 hours. They can be exchanged concurrently during normal operation.

To provide high data availability, RAMAC 1 and RAMAC 2 employed RAID 5 drawer design. RAMAC 3 uses the third generation RAID 5. If one disk fails, you can continue to access the data stored in the drawer. The storage controller can automatically copy the data from the failed drawer to the spare drawer. This function is called *dynamic sparing*. The drawer can perform data reconstruction on a failed disk drive when dynamic sparing has not been performed. You can choose the RAID 5 recovery procedures. The RAID 5 disk array consists of four Ultrastar 2XP 9.1 GB disk drives.

A.3.2.2 Remote Copy

Remote Copy increases data availability in the event of disaster. Remote Copy automatically manages secondary user data at the remote location. Remote Copy is a storage-based disaster recovery and workload migration solution that provides the capability to copy data in real time to a remote location.

Remote Copy offers two options for your disaster recovery and workload migration needs:

- "Extended Remote Copy" on page 375
- "Peer-to-Peer Remote Copy" on page 376

Both options address the problem of unrecoverable data that occurs between the last, safe backup of an application system to a recovery system and the time when the application system fails.

Extended Remote Copy

Extended Remote Copy (XRC) is a solution based on hardware and software for disaster recovery and workload migration management. XRC operates with a combination of 3990 or 9390 Licensed Internal Code (LIC) and DFSMS/MVS software. Data specified at a volume level can be copied asynchronously from a supported 3990 or 9390 at the primary site to a 3990, 9390, or 3390-like geometry subsystem (for example RVA Turbo X83) at a recovery site. The primary storage control can be attached to the XRC DFSMS/MVS host at up to 43 km through ESCON links, or at any distance through the network by mean of converter boxes from ESCON protocol to network and vice versa protecting your configuration from any regional disasters. See your IBM marketing representative for the most current list of link extension suppliers. Channel extension products that use protocol conversion can be used to provide distances of thousands of miles.

Depending on the overall system workload, and because XRC is an asynchronous copy solution, there may be little or no performance impact to the primary site DASD subsystem due to the use of XRC. The primary storage control's DASD writes take place normally, with no delay. XRC DFSMS/MVS data mover periodically reads those writes and subsequently transfers them to the recovery site DASD.

The DFSMS/MVS System Data Mover (SDM) maintains a strict sequence of updates to the recovery site storage control to protect against data corruption. Queries to the primary site storage control side file are frequent enough (at customer-selected time intervals, and Storage Control request) to ensure that, in the event of an outage at the primary site, only a small number of updates could be lost between the two sites.

See *Planning for IBM Remote Copy*, SG24-2595 for more details.

XRC is illustrated in Figure 122.
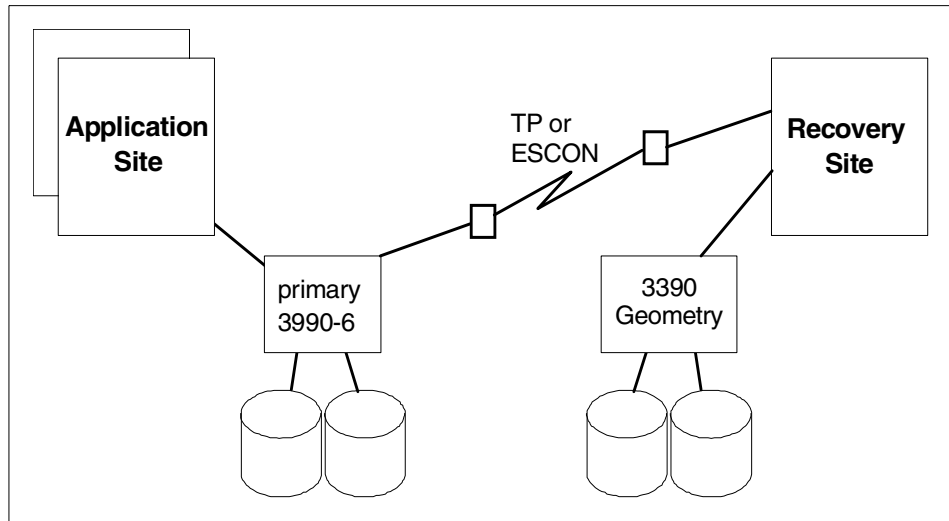
Figure 122. Extended Remote Copy

Support storage subsystems

XRC application site storage controls must be:

- 3990-6, 9390-001, or 9390-002

XRC recovery site storage control can be:

- 3990-6, 9390-001, or 9390-002
- 3990-3, RAMAC Array DASD, RAMAC Virtual Array, or RAMAC Scalable Array, or any 3390-like geometry subsystem.

IBM recommends that the XRC recovery site storage control be a 3990-6, 9390-001, or 9390-002 because of their functional and performance attributes. However, the 9390-002 offers a unique capability that improves XRC performance and makes it a candidate for recovery site storage control.

Peer-to-Peer Remote Copy

Peer-to-Peer Remote Copy (PPRC) is a hardware solution to the problem of accurate and rapid disaster recovery, and also provides a workload migration and DASD migration solution.

Peer-to-peer remote copy is designed for those sites that:

- Need the recovery system to always be fully current with the primary application system.
- Can accept some performance impact to application write I/O operations at the primary location.

A PPRC data copy to the recovery storage subsystem is synchronous with the primary volume's I/O operation. This means that when the application system writes data to a primary volume's storage subsystem DASD, and this data has been transferred to cache and nonvolatile storage (NVS), the storage control sends channel end status to the application host system channel. The application site storage control then initiates an I/O channel program to the recovery site storage control to transfer the updated data.

The application site storage control returns device-end status to the application system when the transfer to the recovery-site storage control cache and NVS is complete. The application system then notifies the application program that the operation is complete.

The PPRC copy function does not consider the application-system DASD write operation complete until the data sent to the recovery storage subsystem has received channel-end and device-end status from the secondary volume's storage control. Each application system write to the recovery subsystem causes an increase to the application system response time.

The primary storage control can be attached to the secondary at up to 43 km through ESCON links.

See *Planning for IBM Remote Copy*, SG24-2595 for more details.

Peer-to-peer remote copy dynamic address switching, added to PPRC, (P/DAS), improves the availability of your data during migration. This enables you to migrate individual volumes or entire subsystems with minimal interruption to your system, application, or users, while simultaneously freeing up resources and improving productivity across your entire enterprise.

See *P/DAS and Enhancements to the IBM 3990-6 and RAMAC Array Family*, SG24-4724 for more details.

PPRC is illustrated in Figure 123.
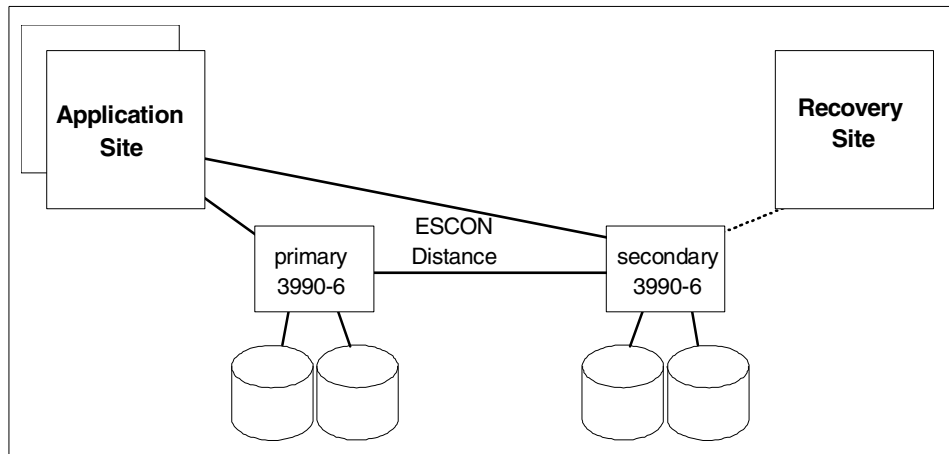
Figure 123. Peer-to-Peer Remote Copy: remote recovery site example

Support storage subsystems
PPRC is available for a combination of 3990-6 and 9390, and for RVA Turbo X83.

A.3.2.3 Advanced cache management
RAMAC 3 array storage has advanced functions that provide efficient cache managements.

- Sequential detect — With the sequential detect function, applications that process data sequentially, without setting the appropriate software hint for sequential access, can take advantage of sequential prestaging. For applications such as VSAM KSDS processing, and other database applications, this function reduces elapsed time.

- Record cache — The record cache function for read I/Os stages only the record or records requested into cache. This provides for more economical use of the cache. Staging only the requested record or records frees cache space for other activity. This increases the probability that the data will not be destaged before being needed again by the application, thereby improving overall performance.

A.3.2.4 Concurrent Copy
Concurrent Copy is a 3990 or 9390 extended function designed to facilitate the copying of important data by significantly reducing the time during which applications cannot update data that is being backed up (reducing it from possibly hours of planned unavailability, to just minutes, in most cases).

Concurrent Copy can increase the scheduling flexibility of on-line operations and batch processing in an OS/390 environment.

The Concurrent Copy function enables you to take *point-in-time* copies of data by using a cache sidefile that is loaded with the *time-zero* version of the data. *Time-zero* refers to the state of the data when the Concurrent Copy session is started and before it gets updated. Before a record is updated, it is copied to the cache sidefile, thus creating a *before update* version or *time-zero* version of the record. After the cache sidefile data has been prepared, the record can be updated on the physical device — the physical device now contains the updated version while the cache sidefile contains the time-zero version of the data.

The Concurrent Copy function is illustrated in Figure 124.
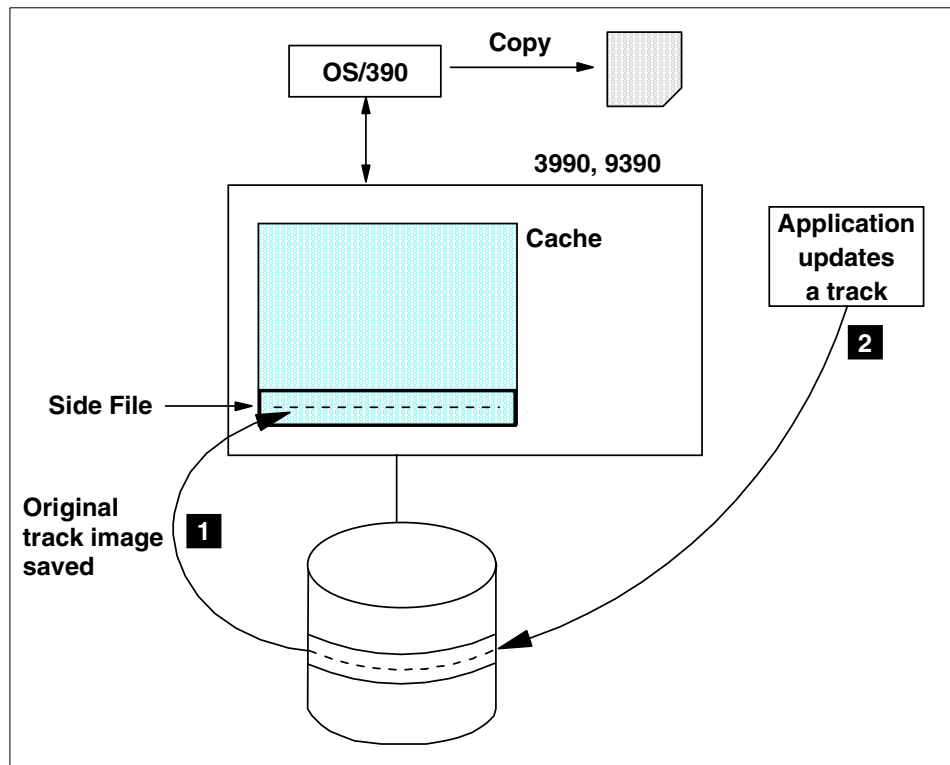
Figure 124. Concurrent Copy

Concurrent Copy appears to the application to be instantaneous, but since a physical copy is actually recorded, it does take the amount of time needed to actually complete the copy.

Activation of Concurrent Copy can be invoked by either DFSMSdss or by DFSMShsm functions. Some procedural changes may be required for Concurrent Copy of database data (such as IMS, DB2, and CICS).

With DFSMSdss, there is the option of using Concurrent Copy when data sets are copied. With DFSMShsm, system-managed data may be backed up automatically with Concurrent Copy by use of management class attributes. DFSMShsm also allows Concurrent Copy to be used when copying data using Aggregate Backup and Recovery Support (ABARS).

See 13.2, "Availability management" on page 256, and *Implementing Concurrent Copy,* GG24-3990, for more details.

The comparison of Concurrent Copy backup process with the traditional backup process is illustrated in Figure 125.
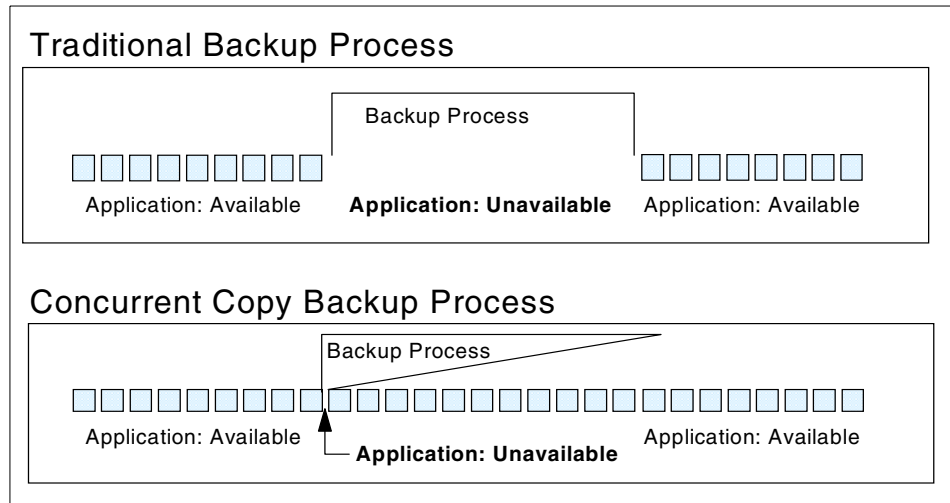
Figure 125. Concurrent Copy backup process

Support storage subsystems
3990 Model 6, 3990 Model 3, or 9390 are required for Concurrent Copy function.

A.3.2.5 3990-6 investment protection

The existing 3990 Model 6 Storage Control Unit can be used as a RAMAC 3 storage control unit by simply implementing a microcode upgrade. The 3990-6/RAMAC 3 combination supports up to 128 device address, and up to 363 GB in a single RAMAC3 rack. The 3990-6/RAMAC3 combination supports ESCON and parallel channel attachments.

A.3.3 Specifications

The specifications of RAMAC 3 are as follows:

- Storage capacity (min/max): 45.4/363 GB (9390-001), 45.4/726 GB (9390-002)

- Cache capacity (min/max): 256/4096 MB (9390-001), 512/8092 MB (9390-002)

- NVS capacity (min/max): 32/128 MB (9390-001), 64/256 MB (9390-002)

- ESCON Channels (min/max/logical paths): 4/16/128 (9390-001), 8/32/256 (9390-002)

- 9391-A30 Storage Frame: Storage capacity (min/max): 45.4/363 GB

- 9392 B33 Drawer: Storage capacity: 22.7 GB

Minimum software requirements

The minimum software needed is as follows:

- OS/390 Release 1 or MVS/ESA SP Version 3, Release 1.3, and either MVS/DFP Version 3, Release 1.1 or DFSMS/MVS Version 1, Release 1

- VM/ESA Version 1 Release 2.2

- VSE/ESA Version 1 Release 3.0

Appendix B. Special notices

This publication is intended to help storage administrators to implement IBM's storage solutions to establish their own enterprise storage management. The information in this publication is not intended as the specification of any programming interfaces that are provided by individual products explained in this book. See the PUBLICATIONS section of the IBM Programming Announcement for each product for more information about what publications are considered to be product documentation.

References in this publication to IBM products, programs or services do not imply that IBM intends to make these available in all countries in which IBM operates. Any reference to an IBM product, program, or service is not intended to state or imply that only IBM's product, program, or service may be used. Any functionally equivalent program that does not infringe any of IBM's intellectual property rights may be used instead of the IBM product, program or service.

Information in this book was developed in conjunction with use of the equipment specified, and is limited in application to those specific hardware and software products and levels.

IBM may have patents or pending patent applications covering subject matter in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to the IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact IBM Corporation, Dept. 600A, Mail Drop 1329, Somers, NY 10589 USA.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The information contained in this document has not been submitted to any formal IBM test and is distributed AS IS. The use of this information or the implementation of any of these techniques is a customer responsibility and depends on the customer's ability to evaluate and integrate them into the customer's operational environment. While each item may have been reviewed by IBM for accuracy in a specific situation, there is no guarantee that the same or similar results will be obtained elsewhere. Customers

attempting to adapt these techniques to their own environments do so at their own risk.

Any pointers in this publication to external Web sites are provided for convenience only and do not in any manner serve as an endorsement of these Web sites.

The following terms are trademarks of the International Business Machines Corporation in the United States and/or other countries:

e (logo)®	Redbooks
IBM ®	Redbooks Logo

ACF/VTAM	ADSTAR
AIX	AS/400
BookMaster	CICS/ESA
Common User Access	CICS
DATABASE 2	CICS/MVS
DFSMS	CUA
DFSMSdfp	DB2
DFSMShsm	DFSMS/MVS
DFSORT	DFSMSdss
Enterprise Systems Architecture/390	DFSMSrmm
Enterprise Storage Server	FICON
ES/9000	Enterprise System/9000
ESCON	Enterprise Systems Connection Architecture
ESA/390	IBM
IMS	Magstar
MVS/ESA	MVS/DFP
Netfinity	OS/400
OS/390	Parallel Sysplex
RACF	RMF
RAMAC	S/390
StorWatch	Seascape
ServeRAID	Ultrastar
System/390	VM/ESA

The following terms are trademarks of other companies:

Tivoli, Manage. Anything. Anywhere.,The Power To Manage., Anything. Anywhere.,TME, NetView, Cross-Site, Tivoli Ready, Tivoli Certified, Planet Tivoli, and Tivoli Enterprise are trademarks or registered trademarks of Tivoli Systems Inc., an IBM company, in the United States, other countries, or both. In Denmark, Tivoli is a trademark licensed from Kjøbenhavns Sommer - Tivoli A/S.

C-bus is a trademark of Corollary, Inc. in the United States and/or other countries.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Sun Microsystems, Inc. in the United States and/or other countries.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States and/or other countries.

PC Direct is a trademark of Ziff Communications Company in the United States and/or other countries and is used by IBM Corporation under license.

ActionMedia, LANDesk, MMX, Pentium and ProShare are trademarks of Intel Corporation in the United States and/or other countries.

UNIX is a registered trademark in the United States and other countries licensed exclusively through The Open Group.

SET, SET Secure Electronic Transaction, and the SET Logo are trademarks owned by SET Secure Electronic Transaction LLC.

Other company, product, and service names may be trademarks or service marks of others.

Appendix C. Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this redbook.

C.1 IBM Redbooks

For information on ordering these publications see "How to get IBM Redbooks" on page 393.

IBM Enterprise Storage and Systems Management Solutions

Cross Platform Solutions

- *Data Sharing: Implementing InfoSpeed Copy-Sharing Solutions Among Heterogeneous Systems*, SG24-4120
- *Storage Networking Virtualization: What's it all about?,* SG24-6210

Enterprise Storage Software

- *Principles of Enterprise Storage Management*, SG24-5250
- *DFSMShsm Primer*, SG24-5272
- *DFSMS Optimizer: The New HSM Monitor/Tuner*, SG24-5248
- *DFSMShsm ABARS and Mainstar Solutions*, SG24-5089
- *DFSMS Optimizer Usage Guide*, SG24-2235
- *DFSMS/MVS 1.5 Presentation Guide*, SG24-5356
- *Get DFSMS FIT: Fast Implementation Techniques*, SG24-2568
- *DFSMS FIT: Fast Implementation Techniques Installation Examples*, SG24-2569
- *DFSMS FIT: Fast Implementation Techniques Forms and Foils*, SG24-2570
- *Introduction to SAN*, SG24-5470
- *Data Sharing: Heterogeneous Data Copy Sharing Implementation Guide*, SG24-4120
- *Implementing SnapShot*, SG24-2241
- *Planning for IBM Remote Copy*, SG24-2595
- *P/DAS and Enhancements to the IBM 3990-6 and RAMAC Array Family*, SG24-4724
- *TCP/IP Tutorial and Technical Overview*, GG24-3376

387

- *Administering DCE and DFS 2.1 for AIX*, SG24-4714

- *GPFS: A Parallel File System*, SG24-5165

- *OS/390 MVS Parallel Sysplex Capacity Planning*, SG24-4680

- *AIX V4.1 Migration Volume 2*, SG24-4653

- *AIX Storage Management*, GG24-4484

- *AIX Storage Management Products*, GG24-4495

Open Storage Disk

- *IBM Enterprise Storage Server*, SG24-5465

- *Configuring and Implementing the IBM Fibre Channel RAID Storage Server*, SG24-5414

- *Exploiting SSA Disk Subsystems in Sun Solaris Platform Environments*, SG24-5083

- *Monitoring and Managing IBM SSA Disk Subsystems*, SG24-5251

- *Practical Guide to Serial Storage Architecture for AIX*, SG24-4599

- *IBM Versatile Storage Server*, SG24-2221

- *Storage Networking Virtualization: What's it all about?,* SG24-6210

- *Implementing ESS Copy Services on UNIX and Windows NT/2000,* SG24-5757

- *Implementing the Enterprise Storage Server in Your Environment,* SG24-5420

S/390 Disk

- *Implementing ESS Copy Services on S/390,* SG24-5680

- *IBM Enterprise Storage Server,* SG24-5465

- *RAMAC Virtual Array, Peer-to-Peer Remote Copy, and IXFP/SnapShot for VSE/ESA*, SG24-5360

- *RAMAC Virtual Array: Implementing Peer-to-Peer Remote Copy*, SG24-5338

- *Using RVA and SnapShot for BI with OS/390 and DB2*, SG24-5333

- *Implementing DFSMSdss SnapShot and Virtual Concurrent Copy*, SG24-5268

- *RAMAC Virtual Array*, SG24-4951

- *IBM RAMAC 3 Array Storage*, SG24-4835

- *Implementing the Enterprise Storage Server in Your Environment,* SG24-5420

Tivoli Storage Management

- *Getting Started with ADSM: A Practical Implementation Guide*, SG24-5416

- *ADSM Version 3 Technical Guide*, SG24-2236

- *Using ADSM to Back Up and Recover Microsoft Exchange Server*, SG24-5266

- *Using ADSM to Back Up Lotus Notes*, SG24-4534

- *Windows NT Backup and Recovery with ADSM*, SG24-2231

- *ADSM Concepts*, SG24-4877

- *ADSM Server-to-Server Implementation and Operation*, SG24-5244

Removable Media Storage Solutions

- *Enhanced IBM Magstar Virtual Tape Server: Implementation Guide*, SG24-2229

- *IBM Magstar 3494 Tape Libraries: A Practical Guide*, SG24-4632

- *Magstar MP 3575 Tape Library Dataserver: Multiplatform Implementation*, SG24-4983

- *Guide to Sharing and Partitioning IBM Tape Library Data*, SG24-4409

- *Magstar and IBM 3590 High Performance Tape Subsystem Technical Guide*, SG24-2506

- *Magstar and IBM 3590 High Performance Tape Susbsystem: Multiplatform Implementation*, SG24-2594

- *The IBM Magstar Virtual Tape Server and Enhancements to Magstar: New Era in Tape*, SG24-4917

- *Tivoli Storage Management Version 3 Technical Guide*, SG24-5477

- *Implementing Concurrent Copy,* GG24-3990

- *Using ADSM Hierarchical Storage Management*, SG24-4631

- *Using ADSM to Back Up Databases*, SG24-4335

C.2 IBM Redbooks collections

Redbooks are also available on the following CD-ROMs. Click the CD-ROMs button at `ibm.com`/redbooks for information about all the CD-ROMs offered, updates and formats.

CD-ROM Title	Collection Kit Number
IBM System/390 Redbooks Collection	SK2T-2177
IBM Networking Redbooks Collection	SK2T-6022
IBM Transaction Processing and Data Management Redbooks Collection	SK2T-8038
IBM Lotus Redbooks Collection	SK2T-8039
Tivoli Redbooks Collection	SK2T-8044
IBM AS/400 Redbooks Collection	SK2T-2849
IBM Netfinity Hardware and Software Redbooks Collection	SK2T-8046
IBM RS/6000 Redbooks Collection	SK2T-8043
IBM Application Development Redbooks Collection	SK2T-8037
IBM Enterprise Storage and Systems Management Solutions	SK3T-3694

C.3 Other resources

These publications are also relevant as further information sources:

- *3990/9390 Planning, Installation, and Storage Administration Guide*, GA32-0100

- *3990/9390 Introduction*, GA32-0098

- *DFSMS/MVS V1 Remote Copy Guide and Reference*, SC35-0169

- *DFSMS/MVS V1R3 NFS Customization and Operation,* SC26-7029

- *DFSMS/MVS V1R3 NFS User's Guide*, SC26-7028

- *ADSM V3 MVS Admin Guide*, GC35-0277

- *ADSM V3R1 UNIX Hierarchical Storage Management,* SH26-4083

- *MVS/ESA SML: Leading a Storage Administration Group*, SC26-3126

- *DFSMS V1.3 Using ISMF*, SC26-4911

- *DFSMS/MVS V1R2 Using the Volume Mount Analyzer*, SC26-4925

- *DFSMS1.4 OAM PISA for Tape Libraries*, SC26-3051

- *IBM RAMAC 9393 General Information Manual*, GC26-7167

- *ADSM for MVS DRM Administrator's Guide and Reference*, GC35-0238

- *DFSMS/MVS V1R5 DFSMShsm Storage Administration Guide*, SH21-1076

- *DFSMS/MVS DFSMSrmm Guide and Reference*, SC26-4931
- *ADSM Disaster Recovery Manager: Stand-alone Disk Image Dump and Restore Guide*, GC35-0343

C.4 Referenced Web sites

These Web sites are also relevant as further information sources:

- http://www.storage.ibm.com/firsts/timeline.htm
- http://www.storage.ibm.com/hardsoft/products/ess/supserver.htm
- http://www.ibm.com/storage/mss
- http://www.storage.ibm.com/hardsoft/products/fast200/fast200.htm
- http://www.ibm.com/storage/SANGateway
- http://www.storage.ibm.com/hardsoft/products/fchub/fcshub.htm
- http://www.ibm.com/storage/director
- http://www.storage.ibm.com/ibmsan/products/directors/inrange_dir.htm
- http://www.vicom.com
- http://www.ibm.com/storage/FCSwitch
- http://www.storage.ibm.com/storage/software/adsm
- http://www.storage.ibm.com/software/adsm/adsercli.htm
- http://www.storage.ibm.com/storage/software/sms/smshome.htm
- http://www.ibm.com/services
- http://www.ibm.com/services/learning
- http://www.ibm.com/storage/fcss

How to get IBM Redbooks

This section explains how both customers and IBM employees can find out about IBM Redbooks, redpieces, and CD-ROMs. A form for ordering books and CD-ROMs by fax or e-mail is also provided.

- **Redbooks Web Site** `ibm.com`/redbooks

  Search for, view, download, or order hardcopy/CD-ROM Redbooks from the Redbooks Web site. Also read redpieces and download additional materials (code samples or diskette/CD-ROM images) from this Redbooks site.

  Redpieces are Redbooks in progress; not all Redbooks become redpieces and sometimes just a few chapters will be published this way. The intent is to get the information out much quicker than the formal publishing process allows.

- **E-mail Orders**

  Send orders by e-mail including information from the IBM Redbooks fax order form to:

	e-mail address
In United States or Canada	pubscan@us.ibm.com
Outside North America	Contact information is in the "How to Order" section at this site: http://www.elink.ibmlink.ibm.com/pbl/pbl

- **Telephone Orders**

United States (toll free)	1-800-879-2755
Canada (toll free)	1-800-IBM-4YOU
Outside North America	Country coordinator phone number is in the "How to Order" section at this site: http://www.elink.ibmlink.ibm.com/pbl/pbl

- **Fax Orders**

United States (toll free)	1-800-445-9269
Canada	1-403-267-4455
Outside North America	Fax phone number is in the "How to Order" section at this site: http://www.elink.ibmlink.ibm.com/pbl/pbl

This information was current at the time of publication, but is continually subject to change. The latest information may be found at the Redbooks Web site.

IBM Intranet for Employees

IBM employees may register for information on workshops, residencies, and Redbooks by accessing the IBM Intranet Web site at http://w3.itso.ibm.com/ and clicking the ITSO Mailing List button. Look in the Materials repository for workshops, presentations, papers, and Web pages developed and written by the ITSO technical professionals; click the Additional Materials button. Employees may access MyNews at http://w3.ibm.com/ for redbook, residency, and workshop announcements.

IBM Redbooks fax order form

Please send me the following:

Title	Order Number	Quantity

First name _____ Last name _____

Company _____

Address _____

City _____ Postal code _____ Country _____

Telephone number _____ Telefax number _____ VAT number _____

☐ Invoice to customer number _____

☐ Credit card number _____

Credit card expiration date _____ Card issued to _____ Signature _____

We accept American Express, Diners, Eurocard, Master Card, and Visa. Payment by credit card not available in all countries. Signature mandatory for credit card payment.

Abbreviations and acronyms

APA	all points addressable		**FC–AL**	fiber-channel arbitrated loop
IBM	International Business Machines Corporation		**FC–PH**	fiber channel physical and signaling interface
ITSO	International Technical Support Organization		**FC-RSS**	Fiber Channel RAID Storage Server
ABARS	aggregate backup and recovery support		**FCS**	Fiber Channel Standard
ACF	Automatic Cartridge Facility		**FICON**	Fiber Connection
ADSM	ADSTAR Distributed Storage Manager		**FOR**	file-owning regions
			GMR	giant magnetoresistive
ATM/AAL	asynchronous transfer method/ATM adaptation layer		**HSM**	hierarchical storage management
CCW	continuous composite WORM		**IND**	interactive network dispatcher
CMOS	complementary metal oxide semiconductor		**IXFP**	IBM eXtended Facility Product
COLD	computer output to laser disk		**LSF**	log structured file
CRC	cyclic redundancy checking		**MR**	magnetoresistive
			NLM	network lock manager
CTQ	command tag queuing		**NSM**	network status manager
DRM	disaster recovery management		**NVS**	nonvolatile storage
ECC	error correction code		**OEMI**	original equipment manufacturer interface
EDAP	Extended Data Availability and Protection		**PRML**	partial response maximum likelihood
ESA	Enterprise Systems Architecture		**RAB**	RAID Advisory Board
ESCON	Enterprise System Connection		**RAID**	Redundant Array of Independent Disks
ESS	Enterprise Storage Server		**RAMAC**	Random Access Method of Accounting and Control
FC	fiber channel		**SAN**	storage area network
			SCSI	small computer systems interface

SPI	SCSI-3 Parallel Interface
SSA	serial storage architecture
SSAIA	Serial Storage Architecture Industry Association
SSI	subsystem interface
TCP/IP	transmission control protocol/Internet protocol
ULP	upper level protocols
VSS	Versatile Storage Server
WCM	Web cache manager
WORM	write once, read many
WTE	Web Traffic Express

Index

Numerics

409

IBM Redbooks review

Your feedback is valued by the Redbook authors. In particular we are interested in situations where a Redbook "made the difference" in a task or problem you encountered. Using one of the following methods, **please review the Redbook, addressing value, subject matter, structure, depth and quality as appropriate.**

- Use the online **Contact us** review redbook form found at **ibm.com**/redbooks
- Fax this form to: USA International Access Code + 1 845 432 8264
- Send your comments in an Internet note to redbook@us.ibm.com

Document Number **Redbook Title**	SG24-5250-02 Enterprise Storage Solutions handbook
Review	
What other subjects would you like to see IBM Redbooks address?	
Please rate your overall satisfaction:	O Very Good O Good O Average O Poor
Please identify yourself as belonging to one of the following groups:	O Customer O Business Partner O Solution Developer O IBM, Lotus or Tivoli Employee O None of the above
Your email address: The data you provide here may be used to provide you with information from IBM or our business partners about our products, services or activities.	O Please do not use the information collected here for future marketing or promotional contacts or other communications beyond the scope of this transaction.
Questions about IBM's privacy policy?	The following link explains how we protect your personal information. **ibm.com**/privacy/yourprivacy/

IBM

Redbooks

Enterprise Storage Solutions Handbook

(0.5" spine)
0.475"<->0.875"
250 <-> 459 pages

Enterprise Storage Solutions Handbook

IBM ®

Redbooks

Confused about which storage solution you should use?

Read this book for all (or some) of the answers

Learn about products, features, and solutions

Mission-critical data to run your e-business is being distributed from enterprise servers to functional UNIX servers, departmental Windows NT servers, and your desktop PCs, to get the maximum benefit out of each platform. To avoid data loss, the storage resources holding that data must be well managed. For example, IBM mainframes such as S/390 systems offer system managed storage (SMS) capability.

This IBM Redbook is intended to help you understand what enterprise storage management is; the key technologies to know; the IBM subsystems, software, and solutions available today; and how to implement various enterprise storage administration tasks, so that you can establish your own enterprise storage management environment.

The book covers four main topics: Introduction to Enterprise Storage Management, Understanding Storage Subsystems, Understanding Storage Management Software, and Implementing Enterprise Storage Administration. We start by introducing the Seascape architecture, Storage Area Network (SAN) Initiative, and Data Sharing solutions. We discuss key technologies used in storage subsystems today, including IBM storage subsystems currently available for both Open Systems and S/390 platforms. Finally, we provide overviews of the IBM storage management software, such as Tivoli Storage Manager, DFSMS/MVS, and StorWatch.